Thermodynamic modeling explains the regulation of *CYP1A1* expression in the liver

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

d o c t o r r e r u m n a t u r a l i u m
(Dr. rer. nat.)
im Fach Biophysik

eingereicht an der
Lebenswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
Dipl.-Ing. Pascal Schulthess

Präsident der der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Lebenswissenschaftlichen Fakultät:
Prof. Dr. Richard Lucius

Gutachter:
1. Prof. Dr. Nils Blüthgen
2. Dr. Stefan Legewie
3. Prof. Dr. Michael Schwarz

Tag der mündlichen Prüfung: 04.03.2016

Abstract

On a daily basis our body is exposed to countless foreign and potentially harmful substances. When ingested such xenobiotics are passed to the liver where they are biotransformed by enzymes such as cytochrome P450, family 1, subfamily A, polypeptide 1 (CYP1A1) into less harmful and/or easily excretable metabolites. Such enzymes are often zonally expressed in the lobes of the liver because of the regiospecific mode of action of xenobiotics as well as the spatiotemporal distribution of oxygen and nutrients. The expression of the *CYP1A1* gene for instance is restricted to the central region of liver lobules and widens in direction of the portal field following exposure to the toxin 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD). This zonation arises, among other factors, in response to the interconnection of two signal transduction pathways, namely AhR and Wnt/β-catenin signaling. The underlying mechanisms by which these two signal transduction pathways orchestrate zonated expression as well as the influence of the *cis*-regulatory region of the *CYP1A1* gene are, however, still poorly understood.

The study at hand presents an analysis of the integration of the AhR and the Wnt/β-catenin signaling pathways into the *CYP1A1* promoter as well as the regulatory influence of the promoter logic on gene expression. Experimentally, this analysis was conducted with the help of 29 mutant constructs of the human *CYP1A1* promoter. I complemented this experimental approach with a set of mathematical models that combined a representation of the signaling crosstalk with a statistical mechanics description of the combinatorial promoter occupancy. With the help of well controllable synthetic promoter constructs I found that only the dioxin responsive element closest to the transcription start site communicates the promoter occupancy to the RNA polymerase. Furthermore, transcription factors only interact with transcription factors that associate with nearby binding sites, i.e., no long-distance binding was observed. The modeling approach subsequently enabled the successful prediction of an AND-gate-like integration of the two signaling pathways into the promoter.

For the genomic architecture of the *CYP1A1* promoter, I could demonstrate the importance of the Wnt/β-catenin pathway target binding site within the *cis*-regulatory region. The model uncovered that this binding site is the strongest

and most promiscuous interaction partner of the remaining transcription factors. In addition, a less switch-like response to the integration of the two signaling pathways as compared to the all-or-none AND-gate within the synthetic constructs could be demonstrated. And lastly, the physiological expression pattern in liver lobules could be successfully predicted by the model and experimentally verified.

In conclusion, in this study I found that crosstalk between AhR and Wnt/β-catenin signaling is crucial for a sensitive regulation of zonated *CYP1A1* expression. Additionally, it exemplifies how statistical mechanics modeling in combination with combinatorial reporter assays has the capacity to disentangle even the complex architectures of eukaryotic promoter systems.

Zusammenfassung

Tagtäglich ist unser Körper unzähligen fremden und potentiell gesundheitsschädlichen Substanzen ausgesetzt. Nachdem solche Xenobiotika über die Nahrung aufgenommen und an die Leber weitergeleitet wurden, beginnen Enzyme wie zum Beispiel Cytochrom P450, Familie 1, Unterfamilie A, Polypeptid 1 (CYP1A1) ihre Biotransformation in weniger schädliche und/oder leichter ausscheidbare Metabolite. Aufgrund der regiospezifischen Wirkungsweise der Xenobiotika sowie der räumlichen und zeitlichen Verteilung von Sauerstoff und Nährstoffen in den Läppchen der Leber, ist die Expression von Enzymen die die Reaktionen der Biotransformation katalysieren oft auch zonal beschränkt. Die Expression des *CYP1A1* Gens ist zum Beispiel auf den Bereich um die Zentralvene eines Leberläppchens beschränkt, wenn keine Exposition durch das Toxin 2,3,7,8- Tetrachlorodibenzo-p-dioxin (TCDD) vorliegt. Nach TCDD Exposition vergrößert sich der Expressionsbereich jedoch in Richtung der portalen Trias. Diese Zonierung ist unter anderem auf die enge Verschaltung des AhR Signaltransduktionswegs mit dem Wnt/β-catenin Signaltransduktionsweg zurückzuführen. Die zugrundeliegenden Mechanismen nach welchen diese Signaltransduktionswege die zonale Expression orchestrieren sowie der Einfluss der *cis*-regulatorischen Region des *CYP1A1* Gens sind jedoch noch immer wenig verstanden.

Die vorliegende Studie präsentiert eine Analyse der Integration der AhR und Wnt/β-catenin Signalwege in den *CYP1A1* Promotor sowie den regulatorischen Einfluss der Promotorlogik auf die Genexpression. Experimentell wurde diese Analyse mithilfe 29 mutagener Reporterkonstrukte des humanen *CYP1A1* Promotors durchgeführt. Ein mathematisches Modell, welches eine Repräsentation des Crosstalks der Signaltransduktionswege mit einer statistisch mechanischen Beschreibung der kombinatorischen Promotorbelegung kombiniert, komplementierte den experimentellen Ansatz. Unter zusätzlicher Zuhilfenahme von gut kontrollierbaren synthetischen Promotorkonstrukten fand ich heraus, dass nur jenes Dioxin-responsive Element das sich am nächsten am Transkriptionsstartpunkt befindet, die Promotorbelegung an die RNA Polymerase kommuniziert. Außerdem beobachtete ich, dass Transkriptionsfaktoren alleine mit Transkriptionsfaktoren interagieren die mit benachbarten Bindestellen assoziieren, d.h. Interaktionen überbrücken keine größeren Entfernungen.

Der Modellierungsansatz ermöglichte zudem die erfolgreiche Vorhersage einer UND-Gatter-ähnlichen Integration der beiden Signalwege in den Promotor.

Für die genomische Architektur des *CYP1A1* Promotors konnte ich die Signifikanz der Zielbindestelle des Wnt/β-catenin Signalwegs innerhalb des *cis*-regulatorischen Region demonstrieren. Mithilfe des Modells fand ich heraus, dass diese Bindestelle am stärksten und vielfältigsten mit den restlichen Transkriptionsfaktoren interagiert. Zusätzliche konnte, im Vergleich zu dem alles-oder-nichts UND-Gatter der synthetischen Konstrukte, eine sehr viel graduellere Antwort auf die Integration der beiden Signalwege aufgezeigt werden. Abschließend wurde das physiologisch zu beobachtende Expressionsmuster von dem Modell vorhergesagt und experimentell validiert.

Zusammenfassend fand ich in der vorliegenden Studie heraus, dass die Vernetzung zwischen den AhR und Wnt/β-catenin Signalwegen ausschlaggebend für die sensitive Regulation der zonierten *CYP1A1* Expression ist. Außerdem veranschaulichte ich, wie statistisch mechanische Modellierung in Kombination mit kombinatorischen Reporterexperimenten die Fähigkeit besitzt, selbst die komplexen Promotorarchitekturen von eukaryotischen Promotorsystemen zu entflechten.

Contents

1 Introduction

1.1 Hepatic detoxification of xenobiotics

On a daily basis our body is exposed to a myriad of foreign substances that can neither be used for the production of energy or the maintenance of structure nor can they be stored without causing harm. Such xenobiotics present in the form of food contaminants, therapeutic agents, and workplace chemicals or environmental compounds. They are primarily ingested but also taken up through inhalation or transdermally. Within the body, they may burden or damage the organism. Thus, efficient mechanisms to transform them into less harmful and/or easily excretable metabolites have developed very early in evolution (Buters, 2008; Kuntz et al., 2006; Oesch and Arand, 1999).

One prominent example of such xenobiotics is 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD). While TCDD has no industrial or commercial use it is often produced as a side-product in organic synthesis (e.g. for the production of pesticides or herbicides), by burning organic material (e.g. in emissions of waste incinerators, forest fires, exhaust emissions, cigarette smoke etc.), or other industrial (e.g. metallurgical) processes. Following emission, TCDD is deposited in the environment on soil, gras, vegetation, and surface water. Thus, while TCDD accumulates preferentially in fish, reptiles, birds and mammals, humans are mainly exposed through the consumption of food such as milk, dairy products, fish, meat and eggs (Koss and Wölfle, 1999). Repeated or highly dosed exposure to TCDD can cause chloracne, progressive liver failure, emphysema, renal failure, myocardial degeneration, and is especially dangerous during development (Stockinger et al., 2014).

The physiological as well as the molecular processes involved in the elimination of TCDD from the body will be introduced in the subsequent chapters.

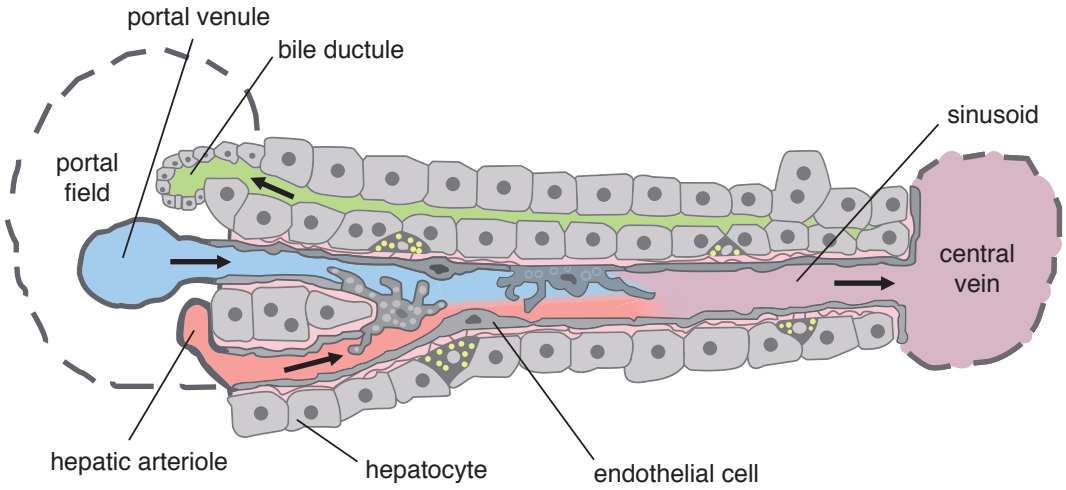


Figure 1.1: Physiological structure of a liver lobule in cross section. The portal field containing portal venule, bile ductule and hepatic arteriole is connected to the central vein by sinusoids. Hepatocytes, endothelial cells and many other cells types are exposed to blood flowing centripedally from the portal to the central area. Bile flows centrifugally towards the portal triad. In the sinusoids nutrient rich blood from the portal venule mixes with oxygen rich blood from the hepatic arteriole. Image taken from Frevert et al. (2005)

1.1.1 Structure and function of the liver

Food-born foreign compounds such as TCDD enter the body via the gastrointestinal tract that is responsible for digestion of food, absorption of nutrients and expulsion of waste. Following oral uptake, chewing, mixing with saliva, the process of peristalsis passes the food down the esophagus to the stomach and the intestines. In the small intestine the gastric contents mix with bile from the liver and digestive enzymes from the pancreas. Substances absorbed by the blood from the small intestine reach the liver via the hepatic portal vein before they enter cardiovascular circulation. In the liver potentially harmful compounds generally are secreted either unchanged or detoxified into bile. About half of the bile reenters the small intestine while the other half is concentrated via the removal of water and stored in the gallbladder. In the small intestine xenobiotics may be reabsorbed. This enterohepatic circulation may trap xenobiotics in the organism (Schwenk, 2008). Apart from detoxification, the liver fulfills numerous other vital functions in the human body such as maintenance of constant amino acid and glucose levels in the blood, synthesis and control of plasma proteins, synthesis of creatine, urea and uric acid, the degradation of porphyrins, and the production of bile, which is essential for the

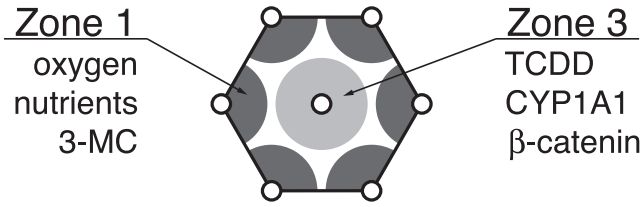


Figure 1.2: Zonation in a liver lobule. Zone 1 around the portal field is rich in oxygen and nutrients while zone 3 is poorly supplied. The toxins 3-MC and TCDD injure zone 1 and 3, respectively. CYP1A1 and β-catenin are mainly found in zone 3.

digestion of fat. The morphological and functional unit of the liver is the hexagon-like structure, called liver lobule. In its corners lie the portal fields consisting of a bile ductule, a portal venule supplying nutrient-rich blood, and an hepatic arteriole through which oxygen-rich blood is supplied (Figure 1.1). Blood from the portal venule and the hepatic arteriole mixes in the sinusoids, drains into the central vein, and enters circulation. Depending on the availability of oxygen and nutrients three zones can be distinguished within a lobule (Figure 1.2). The cells close to the portal field are well provided with oxygen and nutrients (zone 1) while those close to the central vein (zone 3) are poorly provided. The intermediate area is termed zone 2. Resulting from varying levels of oxygen and nutrients the cells in each zone have the ability to perform different metabolic processes and thus contain a different set of enzymes. This biochemical and functional heterogeneity of the cells between the portal and central fields is known as metabolic zonation (Torre et al., 2010). It was discovered that the Wnt/β-catenin signal transduction pathway is a master regulator of zonation (Benhamouche et al., 2006; Colletti et al., 2009; Gebhardt and Matz-Soja, 2014; Gebhardt et al., 2007). Additionally, xenobiotics can cause regiospecific injuries in liver lobules. TCDD, for instance, acts mainly in the centrilobular area while other toxins such as 3-methylcholanthrene (3-MC) injure mainly the portal zone (Gebhardt, 1992). Consequently, enzymes metabolizing toxins are also expressed zonally (Kuntz et al., 2006; Schwarz and Watkins, 2008).

1.1.2 Xenobiotic metabolism and biotransformation

To prevent an accumulation of xenobiotics and thus toxification, gaseous substances are emitted via the lungs, and water-soluble substances are excreted through the

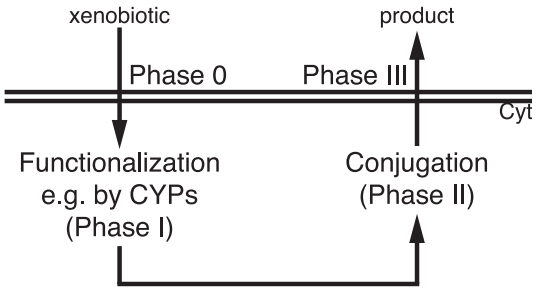


Figure 1.3: Phases of the xenobiotic metabolism. In phase 0 a xenobiotic enters the cell to undergo functionalization in phase I e.g. by cytochrome P450s. Following conjugation in phase II the so produced metabolite is transported back out of the cell in phase III.

kidneys, bile, sweat and intestinal secretions. Lipophilic compounds, which dominate the xenobiotics (70 % – 80 %), on the other hand cannot be excreted in an unchanged state. They are sequestered in body fat and while they can be excreted into the gut or the kidneys, they are rapidly reabsorbed and reenter circulation. To break down or detoxify possibly harmful exogenous or endogenous compounds an efficient mechanism of absorption, distribution, metabolism and excretion is carried out mainly in the hepatocytes. Hepatocytes, which constitute the main parenchymal hepatic tissue not only carry out biotransformation but are also involved in protein synthesis and storage, and the production of bile as a degradation product of cholesterol (Buters, 2008; Kuntz et al., 2006; Oesch and Arand, 1999; Schwarz and Watkins, 2008).

The metabolic processes involved in detoxification are divided into four phases (Figure 1.3). The translocation of the substrate through the cell wall into the cytosol is described in phase 0. For the different forms of xenobiotics this happens in different ways. Lipophilic compounds rapidly reach the liver cells because they are highly fat soluble and can therefore diffuse easily though the phospholipid bilayers of cell membranes. Hydrophilic compounds on the other hand require transport systems to reach the sinusoidal side of the hepatocyte membrane. While hydrophilic xenobiotics can readily be transported out of the cells again, lipophilic substances need to be biotransformed into more hydrophilic derivates to enable a renal or biliary excretion. Within the biotransforming cells, the xenobiotic metabolism mainly takes place in the smooth endoplasmic reticulum (ER), partly also in mitochondria. The functionalization reactions of phase I activate or introduce reactive groupings into the substrate to increase their polarity and reactability to phase

II reactions. The reactions of phase I mainly involve oxidation, reduction, hydrolysis and hydration and are carried out by two groups of enzymes: oxidoreductases and hydrolases. In the conjugation reactions of phase II transferases add a hydrophilic residue to the reactive group introduced in phase I. The resulting products are highly acidic and hydrophilic, biologically inactive, and thus, detoxified and easily excretable. And finally, in phase III, the biotransformed (i.e. highly water-soluble and polar) compounds are actively transported back out of the cell into bile and blood without altering their structure (Buters, 2008; Kuntz et al., 2006; Oesch and Arand, 1999).

Cytochrome P450

The proteins of the cytochrome P450 (CYP) superfamily are the most important oxidoreductases that catalyze phase I of the xenobiotic metabolism. In homo sapiens, all 57 members possess the same heme group but differ in the amino acid sequence of the protein portion of the enzyme. The highest concentration of these enzymes can be found in hepatocytes, amounting to 5 % of total liver protein. CYPs are a prominent example for the broad substrate specificity needed in biotransforming enzymes. The ER membrane-bound enzymes metabolize any electron-donating substrate that is appropriately positioned. Within the lobes of the liver, CYPs are most active in the pericentral region (cf. Figure 1.2) compared to a relative low activity in the periportal area (Buters, 2008; Kuntz et al., 2006).

Cytochrome P450, family 1, subfamily A, polypetide 1 (CYP1A1) is found in almost all mammals and catalyzes similar reactions such as those found in the metabolism of polycyclic aromatic hydrocarbons, a class of toxins containing highly carcinogenic elements (Luch, 2005). In homo sapiens, the gene of CYP1A1 is located on chromosome 15 (genomic coordinates (GRCh37): 15:75 011 882–75 017 950) and is separated by a 23 kb fragment that contains no other open reading frames from the gene of cytochrome P450, family 1, subfamily A, polypetide 2 (CYP1A2). *CYP1A1* and *CYP1A2* are oriented in opposite directions and share a common 5'-flanking region (Corchero et al., 2001). In this 5'-flanking region, Kubota et al. (1991) and Corchero et al. (2001) found various transcription factor binding sites (TFBS's). Apart from binding sites for hepatocyte nuclear factor 1 (HNF1), hepatocyte nuclear factor 3 (HNF3) and hepatocyte nuclear factor 4 (HNF4), they found 13 xenobiotic response elements (XREs), also known as dioxin responsive elements (DREs),

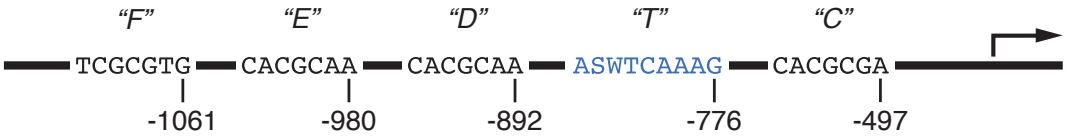


Figure 1.4: Structure of the human *CYP1A1* promoter. The binding site sequences of the DREs are displayed in black and termed *C*, *D*, *E*, and *F*. The binding site sequence of transcription factor 7-like 2 (TCF7L2) is marked in blue and termed *T*. The positions relative to the transcription start site are shown.

on both strands of the DNA. These DREs exhibit `5'-TNGCGTG-3'` as consensus sequence (Swanson et al., 1995; Whitlock et al., 1996; Yao and Denison, 1992). Kress et al. (1998) showed that of the 10 DREs within the 1 400 bp enhancer region of *CYP1A1* only four were functional enhancer sequences able to recruit transcription factors (TFs). In the present study, I followed their terminology, and named them alphabetically from *C* to *F* (Figure 1.4). Additionally, Braeuning et al. (2011) showed that the *CYP1A1* promoter also harbors a binding site for TCF7L2 between the C- and the D-DRE. The consensus sequence of the TCF7L2 binding site is `5'-ASWTCAAAG-3'` (Hatzis et al., 2008).

Upstream of the *CYP1A1* promoter several signal transduction pathways orchestrate the assembly of TFs binding to the *cis*-regulatory region. In the following I will introduce the role of signaling transduction in the liver, and subsequently elaborate on the pathways essential for the control of the *CYP1A1* promoter in more detail.

1.1.3 Signal transduction pathways in the liver

Cells sense and respond to their environment through complex networks of signal transduction pathways. Therefore, signaling pathways are found in every cell of every living organism and their deregulation can have fatal outcomes. External stimuli such as hormones or as it is here the case, xenobiotics, are first sensed by interaction with a cellular component. This signal reception is mostly carried out by specialized cell surface receptors, but membrane-diffusing chemicals can also bind to cytosolic receptors. Through binding and/or activation (e.g. phosphorylation) of intracellular molecules the signal is subsequently transduced to other chemical forms which in turn trigger a cellular response. Cellular responses are manifold and range from the expression of new proteins to proliferation and programmed cell death.

In liver cells various signal transduction pathways collude in order to uphold the function of the organ. Especially well studied are the mitogen-activated protein kinase (MAPK), the Janus kinase (Jak), the transforming growth factor β (Tgfβ), the tumor necrosis factor α (Tnfα) and the different Wnt signal transduction pathways. Through activation by cytokines and growth factors they orchestrate many different mechanisms ranging from the development of the liver as a whole to proliferation, growth, differentiation, adhesion and apoptosis of hepatocytes and other hepatic cells. Furthermore, the aryl hydrocarbon receptor (AhR) links many of these pathways to detoxification which therefore contribute to the overall toxic response (Puga et al., 2009). The canonical Wnt/β-catenin signaling pathway, for example, plays a crucial role in regulating hepatic expression of the detoxifying CYP enzymes (Braeuning et al., 2009; Hailfinger et al., 2006; Loeppen et al., 2005).

AhR signaling pathway

While some toxins act through unspecific destruction of an organism or its constituting cells, many others trigger a specific response that alters the physiological balance of the system. Largely, these specific responses are mediated through a specific sensor that translates toxic exposure into a physiological response. AhR represents such a sensor for polycyclic aromatic hydrocarbons, dioxin-like compounds (e.g. TCDD) and others (Göttlicher, 1999). AhR is a member of the family of basic helix-loop-helix domain TFs and evolutionary conserved from invertebrates onwards (Hahn et al., 1997). It is widely expressed in the body (Gu et al., 2000) and almost exclusively found in the cytosol (Schmidt and Bradfield, 1996).

In the absence of a ligand AhR is coupled in a cytoplasmic complex bound to actin filaments (Figure 1.5). Several chaperones such as heat shock protein 90 (HSP90), chaperone p23, and AhR-interacting protein (AIP), which both interact with AhR and HSP90, are members of this complex (Antonsson et al., 1995; Carver and Bradfield, 1997; Denis et al., 1988; Grenert et al., 1997; Meyer and Perdew, 1999; Meyer et al., 1998; Nair et al., 1996; Perdew, 1988; Perdew and Bradfield, 1996). Through the chaperone complex, AhR is kept in a conformation of high ligand binding affinity (Pongratz et al., 1992). Following diffusion of TCDD through the cell membrane, binding of the ligand leads to dissociation of AhR/HSP90/p23 from the actin filaments and a conformational change in AhR. This allows AhR to expose a nuclear localization signal to importin β and thus translocation into the nucleus (McGuire

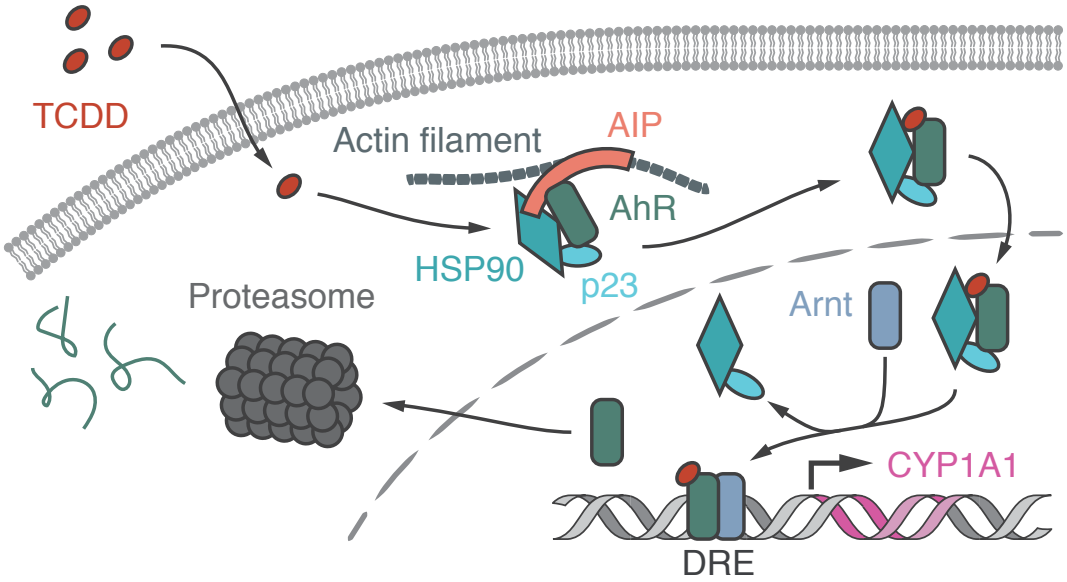


Figure 1.5: AhR signaling pathway. TCDD stimulation releases a complex (containing AIP, HSP90, p23 and AhR) from actin filaments for translocation into the nucleus. There, Arnt induces the heterodimerization with AhR to form a TF that binds to DREs in the promoter region of *CYP1A1*. Subsequently, AhR is degraded by the proteasome.

et al., 1994; Stockinger et al., 2014). Upon binding of AhR nuclear translocator (Arnt) the HSP90-p23 complex is released. Controlled by p23, Arnt can only disrupt the AhR/HSP90 complex following ligand binding and nuclear translocation (Kazlauskas et al., 1999, 2001). This activation renders the heterodimeric complex of AhR and Arnt as a *bona fide* TF able to associate with DREs in the promoter region of genes such as *CYP1A1*. The TCDD/AhR/Arnt TF complex can affect local chromatin structure through interaction with a chromatin remodeling complex (Okino and Whitlock, 1995; Wang and Hankinson, 2002) which increases the accessibility of the promoter to the transcriptional machinery (Koss and Wölfle, 1999). Subsequently, AhR is subject to proteasomal degradation (Davarinos and Pollenz, 1999). Since AhR senses dioxin exposure and thus delivers adaptive and toxic responses it is naturally involved in various toxin-associated liver conditions such as hepatocellular damage and carcinoma (Bock, 1994; Schneider et al., 2014; Stockinger et al., 2014).

The AhR signaling pathway maintains a two-way crosstalk with the Wnt/β-catenin signal transduction pathway (Schneider et al., 2014). On the one hand AhR is a target gene of the Wnt/β-catenin signaling pathway (Chesire et al., 2004) while

AhR activation on the other hand induces a deregulated expression of members of the Wnt/β-catenin signaling pathway (Mathew et al., 2008).

Wnt/β-catenin signaling pathway

The canonical Wnt/β-catenin signaling pathway is involved in a multitude of processes throughout development and adulthood of all metazoan animals (Clevers and Nusse, 2012). Its important functions in stem cell self-renewal and cell proliferation is the reason that deregulation or mutation within the pathway or its components is associated with various diseases including different cancers, diabetes, osteoporosis, coronary artery disease, metabolic syndrome and many more (Anastas and Moon, 2013; Baron and Kneissel, 2013; Cadigan and Peifer, 2009; Clevers, 2006; Klaus and Birchmeier, 2008; MacDonald et al., 2009; Reya and Clevers, 2005).

Without the presence of external Wnt ligands β-catenin, the central mediator of the pathway, undergoes continuous degradation by the proteasomal machinery in order to prevent cytosolic accumulation and activation of target genes (Figure 1.6). Within the so called destruction complex, the tumor suppressor protein axis inhibitor 1 (AXIN1) acts as a scaffold by interacting with β-catenin, the tumor suppressor protein adenomatous polyposis coli (APC) and the two constitutively active serine-threonine kinases casein kinase 1α/δ (CK1α/δ) and glycogen synthase kinase 3α/β (GSK3α/β) (Clevers and Nusse, 2012). These two kinases then sequentially phosphorylate β-catenin (Clevers and Nusse, 2012). The phosphorylated motifs in the N-terminal domain of β-catenin are recognized by the F box/WD protein βTrCP which is part of the E3 ubiquitin ligase complex (Clevers and Nusse, 2012). Subsequently, β-catenin is ubiquitinated and targeted for destruction by the proteasome (Aberle et al., 1997; Clevers and Nusse, 2012). In the nucleus, the transcriptional repressor Groucho interacts with TFs of the TCF/LEF family to prevent β-catenin binding and thus transcription (Cavallo et al., 1998; Roose et al., 1998).

Presence of extracellular Wnt proteins leads to the formation of a membrane-bound heterodimeric receptor complex consisting of Frizzled (FZD) and low-density lipoprotein receptor-related protein 5 or 6 (LRP5/6) (Pinson et al., 2000; Tamai et al., 2000; Wehrli et al., 2000). The FZD proteins are seven-transmembrane receptors that contain large extracellular domains for Wnt binding (Bhanot et al., 1996; Dann et al., 2001; Janda et al., 2012). LRP5/6 on the other hand is a large modular, single transmembrane protein whose extracellular domain contains multiple separate

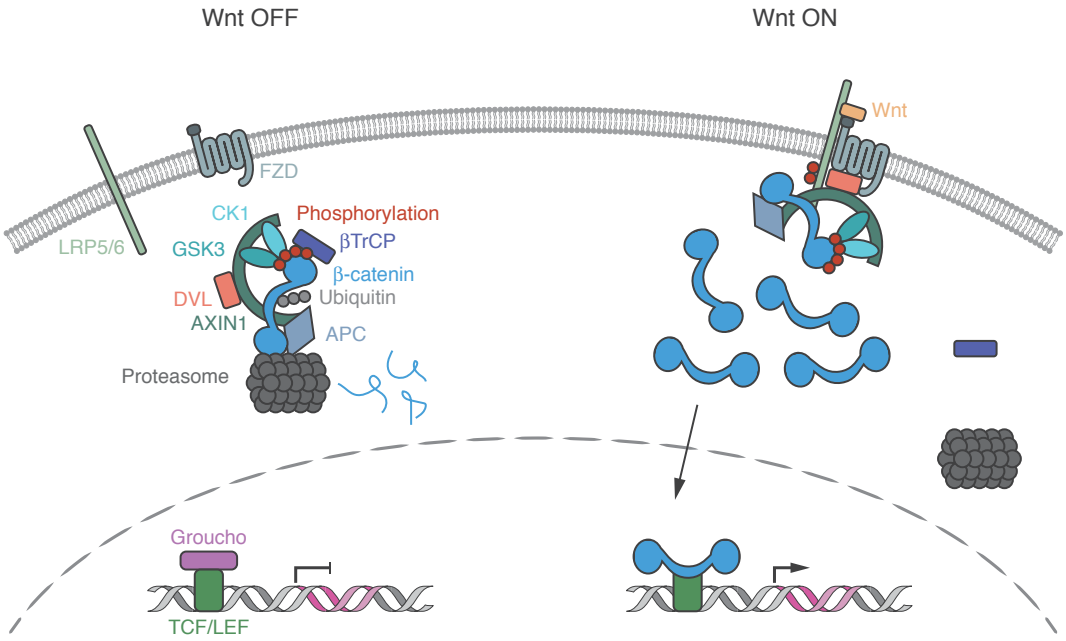


Figure 1.6: Wnt/β-catenin signaling pathway. Following Wnt stimulation β-catenin is released from a destruction complex and can accumulate in the cytosol. After translocation into the nucleus β-catenin associates with TFs such as TCF7L2 to initiate target gene transcription.

Wnt binding sites (Clevers and Nusse, 2012). Together with the promiscuity of the Wnt-FZD interaction (Bhanot et al., 1996) it follows that a single FZD/LRP5/6 can be bound by multiple Wnts and vice versa (Bourhis et al., 2010; Niehrs, 2012). The cytoplasmic part of FZD now interacts with Dishevelled (DVL), which facilitates the recruitment of AXIN1 to the cytoplasmic tail of LRP5/6 (Chen et al., 2011; Mao et al., 2001). The binding between AXIN1 and LRP5/6 is regulated by the phosphorylation of the cytoplasmic tail of LRP5/6 through GSK3α/β and casein kinase 1γ (CK1γ) (He et al., 2004; Tamai et al., 2004). Through this relocalization of the destruction complex to the Wnt-activated receptors the ubiquitination of β-catenin within the complex is inhibited, which leads to a saturation of the complex with phosphorylated β-catenin. Therefore, newly synthesized β-catenin can accumulate in the cytoplasm and translocate to the nucleus with the help of microtubules and active transport (Sugioka et al., 2011). In the nucleus, β-catenin interacts with TFs of the TCF/LEF family including TCF7L2 to initiate the transcription of target genes such as *CYP1A1* (Braeuning et al., 2009; Molenaar et al., 1996).

Having introduced the biology necessary for the comprehension of the present

study, I will now introduce the theoretical methodologies and their applications with which transduction of signals, transcriptional regulation and expression of a gene can be described and analyzed.

1.2 Modeling gene expression and transcriptional regulation

Transcription is the first step towards the expression of a gene. Endogenous or exogenous signals are transduced via signal transduction pathways and lead to the activation of special proteins that possess the ability to bind to DNA. These TFs initiate transcription through binding to *cis*-regulatory elements in DNA such as promoter and enhancer regions. Subsequently, the RNA polymerase (RNAP) is recruited to the promoter region of the gene of interest and proceeds with the assembly of RNA (Lee et al., 2004).

For the quantitative understanding, experimental techniques are often complemented with modeling approaches that tackle different steps of this process. Gene regulatory networks are models describing the whole process of gene regulation including TF activation/complex formation, TF-TF interaction and the effect of gene products on TFs (de Jong, 2002; Karlebach and Shamir, 2008). Recently these networks have been expanded to represent genome wide gene regulation by TFs (MacNeil and Walhout, 2011; Spitz and Furlong, 2012). In the following, I will introduce three important subsets of gene regulatory networks, namely the modeling of TF activation/complex formation by means of signal transduction networks, a bioinformatic approach to describe the specificity and affinity of TFs to their DNA binding sites, and a statistical mechanics method to model the cooperation or competition of TFs following DNA binding.

1.2.1 Modeling signal transduction pathways

Mathematical models of signal transduction pathways in space and time are well established for more than a decade now. Thus, the detailed principles of these models were out of the scope of this study and interested readers are referred to a large body of literature (e.g. Cary et al., 2005; Heinrich and Neel, 2002; Hlavacek et al.,

2006; Ideker et al., 2001; Kitano, 2002a,b,c; Levchenko, 2003; Schulthess and Blüthgen, 2010).

Signal transduction networks are built upon sequences of biochemical reactions whose only cause is the change in concentration. Such reactions can be binding or separation, phosphorylation or as used in this study the TF complex formation (Saez-Rodriguez et al., 2009; Stelniec-Klotz et al., 2012). The temporal behavior of the substances' concentrations $\boldsymbol{x} = (x_i)$ is given by the balance equation

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{N}\boldsymbol{v}\left(\boldsymbol{x}(t), \boldsymbol{p}\right) \ , \tag{1.1}$$

where $\boldsymbol{N} = (n_{ij})$ denotes the stoichiometric matrix describing i substances and j reactions. The reactions rates are summerized in the vector $\boldsymbol{v} = (v_j)$ while $\boldsymbol{p} = (p_k)$ collects the parameters. These chemical reactions can be modeled by means of assuming various rate laws such as mass action or Michaelis-Menten kinetics (Heinrich and Schuster, 1996). The law of mass action, for example, states that substances undergoing a reversible reaction reach a dynamic equilibrium after a certain time, i.e., $v_j = 0$. This leads to

$$k_j^+ \prod_i x_i^{n_{ij}^-} = k_j^- \prod_i x_i^{n_{ij}^+} \tag{1.2}$$

where k_j^+ and k_j^- represent the forward and reverse rate constants while n_{ij}^+ and n_{ij}^- describe the stoichiometric coefficients of the reactants and the products, respectively. The equilibrium constant K_j now is

$$K_j = \frac{k_j^+}{k_j^-} = \frac{\prod_i x_i^{n_{ij}^+}}{\prod_i x_i^{n_{ij}^-}} \ . \tag{1.3}$$

This expression of the equilibrium constant will frequently be used in this study to model biochemical reactions, especially those present in the formation of the DRE-associating TFs. Various other forms and applications of the law of mass action, e.g. irreversible reactions or reactions happening in non-ideal solutions as well as more complex enzyme kinetics, are discussed by Heinrich and Schuster (1996) and are beyond the scope of this study.

Following the formation or activation of TFs through signal transduction networks, their DNA association properties will be introduced in the following.

1.2.2 Specificity of TF-DNA binding events

TFs bind to promoter and enhancer regions in DNA to regulate transcription. These regulatory regions harbor specific TFBS's that are characterized by the binding preference of a TF to a couple of bases in the DNA sequence. The specificity of a TF is therefore defined as its relative binding affinity to a large number of target sequences, i.e., the probability that a TF binds a specific sequence given the competition from all other present and accessible sites in the DNA (Stormo and Zhao, 2010). When T describes the concentration of a TF, S_i the concentration of a binding site, the reversible binding between them is described with

$$T + S_i \underset{k^-}{\overset{k^+}{\leftrightharpoons}} TS_i \tag{1.4}$$

where k^+ and k^- are the rates for the formation of the TF-DNA complex and its dissociation. When the system is in equilibrium, the equilibrium constant, i.e., the association constant is

$$K_i = \frac{k^+}{k^-} = \frac{TS_i}{T \cdot S_i} \,. \tag{1.5}$$

At a specific instance, S_i can be either bound or unbound. The probability that the TF T is bound to the sequence S_i is

$$P(\text{bound}|S_i) = \frac{TS_i}{S_i + TS_i} \,. \tag{1.6}$$

Herein the numerator represents all states in which the TF is bound to the sequence. The denominator summarizes all possible states, i.e., the sequence is bound by the TF plus the sequence is unbound. Using Equation 1.5 I can rewrite the probability to

$$P(\text{bound}|S_i) = \frac{T}{T + \frac{1}{K_i}} = \frac{1}{1 + \frac{1}{K_i T}} \,. \tag{1.7}$$

The specificity now describes the difference in affinity for all potential binding sites, i.e., the ability of one TF to discriminate between different binding sites. By assuming additivity of the positions within a binding site and that the selection of binding sites is proportional to their binding affinity Stormo and Zhao (2010) defined the specificity as

$$\text{Spec} \equiv \sum_{S_i} \frac{K_i(S_i)}{\sum K_i(S_i)} \ln \frac{K_i(S_i)}{\langle K_i(S_i) \rangle} \,. \tag{1.8}$$

While binding affinities were in the past measured one at the time, recent technological advances allow for their high-throughput determination. Such experimental techniques include but are not limited to microfluidics, e.g. mechanically induced trapping of molecular interactions (MITOMI) (Maerkl and Quake, 2007), surface plasmon resonance (SRP) (Campbell and Kim, 2007; Paul et al., 2009; Shumaker-Parry et al., 2004), protein-binding microarrays (Berger et al., 2006) and many more (Stormo and Zhao, 2010). Even though binding affinities can be measured directly with the mentioned experimental techniques, models can be helpful because they are able to average out the noisy nature of experiments. Furthermore, models have the ability to predict novel binding sites as well as the effects of genetic variations (Stormo and Zhao, 2010). The most simple way to model a TFBS is the consensus sequence. To compute the consensus sequence, a set of sequences in which one suspects a TFBS is aligned and for each position the most common nucleotide represents the consensus at said position. For example the consensus sequence `WTNS` represents a set of four-lettered sequences which contain an equal amount of `A`'s and `T`'s at the first position, only `T`'s at the second position, an equal amount of all letters (`A`, `C`, `G` and `T`) at the third position, and and equal amount of `C`'s and `G`'s at the last position. This representation is however only valid for highly specific binding proteins such as restriction enzymes (D'haeseleer, 2006). Since the association between a TF and a binding site is however degenerate, TFs can bind with varying affinity to a site. As a result, position specific weight matrices (PWMs) were introduced that contain scores for every base in a binding site (Staden, 1984; Stormo, 2000; Stormo and Hartzell, 1989; Stormo et al., 1982). Modeling binding specificity with PWMs implicitly assumes independent binding site contributions. Tomovic and Oakeley (2007) however showed that this assumption does not always hold. Thus, more complex models of binding affinity were developed that include dinucleotides or trinucleotides (Mathelier and Wasserman, 2013; Weirauch et al., 2013). Nevertheless, the assumption that each base contributes individually is in most cases a very good approximation of specific DNA-protein interactions (Benos et al., 2002). Furthermore, there exist multiple other complexities in DNA-protein binding that are hardly covered with PWMs. PWMs are usually a result of a large number of sequences where no discrimination was made for high and low affinity binding sites. Additionally, DNA-binding proteins can exhibit various distinct

binding modes that lead to different motifs (Siggers and Gordân, 2014). For example, Elk-1 is able to associate with the DNA as a monomer or as a dimer (Jolma et al., 2013). Despite all these drawbacks, PWMs remain widely used to describe and predict existing and novel TFBS's. In Chapter 2.3, the deduction of the PWM and its relationship to the binding energy is introduced in more detail.

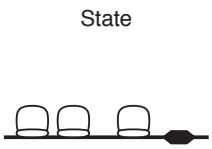
1.2.3 Thermodynamic modeling of *cis*-regulatory elements

Living cells typically act far away from equilibrium. Nevertheless, thermodynamics as a theory of equilibrium can be a versatile tool to solve some biological problems. Since any macroscopic system can be divided in many ways into microscopic objects, i.e., microstates, the thermodynamic view can serve as a flexible framework. Such microstates can for example include all possible arrangements of ligands in a solution or the binding configuration of TFs on a promoter. With the help of statistical mechanics the probability of all microstates under macroscopic constraints such as the concentration of ligands or TFs can be computed. In biological systems the task of finding the probability of all microstates is simplified by assuming constant temperature. Due to the separation of timescales, i.e., temperature changes are much slower than cellular processes, this assumption is reasonable. The probability of all microstates can be calculated with the following formula derived in 1902 by Ludwig Boltzmann

$$p_i = \frac{e^{-\beta\epsilon_i}}{Z_{\text{tot}}} \tag{1.9}$$

where $\beta = 1/k_B T$, k_B is the Boltzmann constant, T is the absolute temperature, ϵ_i is the energy of the corresponding microstate i, and Z_{tot} represents the partition function. To obtain the partition function, the Boltzmann weights $\zeta(i) = e^{-\beta\epsilon_i}$ of all microstates are summed up.

Apart from its use in the description of binding reactions (Dill and Bromberg, 2003; Hill, 1985), the thermodynamic framework is also widely used for models of transcriptional regulation. As one of the earliest examples, Ackers et al. (1982) untangled the regulatory network of *Bacteriophage* λ with the help of statistical thermodynamics. In fact, this framework was mostly applied to simple and experimentally (fully) controllable prokaryotes (Garcia et al., 2010b; Jacob et al., 1960; Johnson et al., 1981; Ptashne and Gann, 2002; Shea and Ackers, 1985). In recent years however the gene expression of increasingly complex organisms, from yeast (Ellis et al., 2009;



State	Energy	Multiplicity	Weight
	$P\epsilon^{NS}$	$\frac{N_{NS}!}{P!(N_{NS}-P)!} \approx \frac{N_{NS}^{P}}{P!}$	$\frac{N_{NS}^{P}}{P!} e^{-\beta P \epsilon^{NS}}$
	$(P-1)\epsilon^{NS} + \epsilon^{S}$	$\frac{N_{NS}!}{(P-1)!(N_{NS}-P+1)!} \approx \frac{N_{NS}^{(P-1)}}{(P-1)!}$	$\frac{N_{NS}^{(P-1)}}{(P-1)!} e^{-\beta((P-1)\epsilon^{NS}+\epsilon^{S})}$

Figure 1.7: Statistical mechanics of RNAP-DNA binding. The depicted model of promoter states asks how many distinct ways exist to distribute P RNAP molecules among N_{NS} non-specific binding sites in the DNA. Two macrostates ("States") show the specific binding site empty (top) or occupied with a RNAP (bottom). All possible ways in which the RNAPs can be distributed across the N_{NS} non-specific sites for both macrostates are counted in the "Multiplicity" column. By assuming ϵ^{NS} and ϵ_S as energies for the RNAP being bound non-specifically and specifically ("Energy"), the weight is the product of the multiplicity and the Boltzmann weight ("Weight"). (Adopted from Garcia et al., 2010a and Bintu et al., 2004).

Gertz and Cohen, 2009; Gertz et al., 2008; Kim and O'Shea, 2008) to worms (Brown et al., 2007), and sea urchins (Yuh et al., 1998) was analyzed with thermodynamic models.

One main application of thermodynamic models is the study of biological cooperativity such as the binding events between receptors and ligands (Garcia et al., 2010a) but applicable to RNAP-DNA or TF-DNA binding events as well. As a basis for these analyses the Monod-Wyman-Changeur (MWC) models are widely used (Monod et al., 1965). The MWC model assumes that individual elements within a complex system can only exist in a countable number of discrete states. Furthermore, each element can sometimes change its state. Since RNAP association to special binding sites within DNA is key to the transcriptional regulation and hence the expression of a gene, I will introduce the application of the thermodynamic framework with the simplest model possible (Figure 1.7). The RNAP-DNA binding system can exist in two macrostates. Either the RNAP binding site in the DNA is unoccupied and the P RNAPs are distributed among N_{NS} non-specific binding sites (Figure 1.7 top row), or one RNAP molecule is bound to its binding site and $P-1$ RNAPs are distributed among the non-specific sites (Figure 1.7 bottom row). However, for each of these macrostates exist multiple different microscopic realizations, i.e., microstates since the RNAPs can be distributed in many different ways among the non-specific sites. The "Multiplicity" column in Figure 1.7 gives an expression

for the number of microstates for each macrostate. In order to find the total statistical weight the multiplicity of the two macrostates has to be multiplied with their associated Boltzmann weights. The probability of either of the two macrostates can be calculated by dividing its statistical weight Z (cf. "Weight" column in Figure 1.7) by the sum of the statistical weights of all possible states, i.e. the partition function Z_{tot}. For the microstate in which the RNAP is bound specifically to its binding site, the probability is

$$p_{\text{bound}}(P) = \frac{Z(P-1)e^{-\beta\epsilon_{PD}^{S}}}{Z(P) + Z(P-1)e^{-\beta\epsilon_{PD}^{S}}} \,. \tag{1.10}$$

Herein, ϵ_{PD}^{S} depicts the binding energy between the RNAP and its specific binding site. By assuming that the number of RNAPs is much less than the number of non-specific sites, the probability that RNAP is bound to DNA is

$$p_{\text{bound}}(P) = \frac{\frac{P}{N_{NS}}e^{-\beta\Delta\epsilon_{PD}}}{1 + \frac{P}{N_{NS}}e^{-\beta\Delta\epsilon_{PD}}} = \frac{1}{1 + \frac{N_{NS}}{P}e^{\beta\Delta\epsilon_{PD}}} \,. \tag{1.11}$$

The loss of energy upon binding to the DNA is $\Delta\epsilon_{PD} = \epsilon_{PD}^{S} - \epsilon_{PD}^{NS}$. In order to express this probability with concentrations, I can rewrite the number of RNAPs $P = [P]N_{NS}$. The probability now takes the form

$$p_{\text{bound}}([P]) = \frac{1}{1 + \frac{K_d}{[P]}} \,, \tag{1.12}$$

with the equilibrium dissociation constant $K_d = e^{\beta\Delta\epsilon_{PD}}$.

The presence of TFs in the *cis*-regulatory region of the promoter can either have activating or repressing effects on the level of gene expression. In order to study this possible regulation, I add an additional activator to the model above. Bintu et al. (2004) listed various ways of regulation including the presence of inducers, multiple activators and/or repressors as well as events of DNA looping. The statistical weight for P RNAP molecules and A activator molecules distributed among N_{NS} non-specific binding sites is given by

$$Z(P, A) = \frac{N_{NS}!}{P!A!(N_{NS} - P - A)!} e^{-\beta(P\epsilon_{PD}^{NS} + A\epsilon_{AD}^{NS})} \,. \tag{1.13}$$

The addition of an activating TF also increases the number of possible macrostates.

Apart from completely empty binding sites and a specifically bound RNAP, there is also the possibility that the activator is bound to a specific site or that the activator and the RNAP specifically bind together. The total statistical weight, i.e., the partition function now takes the form

$$Z_{\text{tot}}(P,A) = \underbrace{Z(P,A)}_{\text{empty sites}} + \underbrace{Z(P-1,A)e^{-\beta\epsilon^S_{PD}}}_{\text{RNAP on promoter}} + \underbrace{Z(P,A-1)e^{-\beta\epsilon^S_{AD}}}_{\text{activator on specific site}} + \underbrace{Z(P-1,A-1)e^{-\beta(\epsilon^S_{PD}+\epsilon^S_{AD}+\epsilon_{PA})}}_{\text{RNAP and activator on specific site}}, \quad (1.14)$$

with ϵ_{AD} and ϵ_{PA} representing the binding energies between the activator and its specific binding site, and the adhesive energy with which the activator recruits the RNAP, respectively. To calculate the probability for the RNAP to be bound to the promoter, i.e., for transcription to be initiated, the corresponding weights are divided by the total statistical weight, such that

$$p_{\text{bound}}(P,A) = \frac{Z(P-1,A)e^{-\beta\epsilon^S_{PD}} + Z(P-1,A-1)e^{-\beta(\epsilon^S_{PD}+\epsilon^S_{AD}+\epsilon_{PA})}}{Z_{\text{tot}}(P,A)}. \quad (1.15)$$

By again assuming $N_{NS} \gg P$ and with the energetic difference between specific and non-specific binding $\Delta\epsilon = \epsilon^S - \epsilon^{NS}$ this expression simplifies to

$$p_{\text{bound}}(P,A) = \frac{1}{1+\frac{N_{NS}}{PF_{\text{reg}}}e^{\beta\Delta\epsilon_{PD}}}. \quad (1.16)$$

Here I introduced the regulation factor

$$F_{\text{reg}} = \frac{1+\frac{A}{N_{NS}}e^{-\beta(\Delta\epsilon_{AD}+\epsilon_{AP})}}{1+\frac{A}{N_{NS}}e^{-\beta\Delta\epsilon_{AD}}} \quad (1.17)$$

that describes the effective change in the number of RNAPs available for promoter binding due to the action of the activator. If $F_{\text{reg}} > 1$ the number of RNAPs increases while $F_{\text{reg}} < 1$ represents a decrease. By using concentrations instead of number of RNAPs and activators, Equations 1.16 transforms into

$$p_{\text{bound}}([P],[A]) = \frac{1}{1+\frac{K_P}{[P]F_{\text{reg}}}}, \quad \text{with} \quad F_{\text{reg}} = \frac{1+\frac{[A]}{K_A}e^{-\beta\epsilon_{AP}}}{1+\frac{[A]}{K_A}} \quad (1.18)$$

where K_P and K_A represent the equilibrium dissociation constants of the RNAP and the activator, respectively. Since it is experimentally often most convenient to measure gene expression fold-changes, the thermodynamic models should also be able to reflect that. For the aforementioned activator model the ratio of the regulated case to the unregulated case is

$$\phi = \frac{p_{\text{bound}}(P, A)}{p_{\text{bound}}(P, A = 0)} \,. \tag{1.19}$$

One of the key assumptions of thermodynamic models is that of thermodynamic equilibrium. And since there are many irreversible steps in the transcriptional process, e.g. the elongation leading to mRNAs, these assumptions need to be well justified. As with the temperature dependence, the separation of time scales plays again an important role. TFs and RNAPs usually bind and dissociate in a much shorter time frame to and from the promoter than those associated with transcription initiation. This means that RNAPs and TFs will have enough time to reach binding equilibrium with the DNA such that mRNAs can be produced from this equilibrium state. In other words, the rate of transcription should depend on the concentrations and the activity of TFs and RNAPs. For NF-κB for example, it was shown that the variation in concentration of nuclear NF-κB is much slower than the binding/release of NF-κB from DNA (Bosisio et al., 2006; Darzacq et al., 2007). Another assumption of thermodynamic models is that the probability of promoter occupancy by RNAPs is linearly proportional to the level of gene expression (Bintu et al., 2004). Clearly this assumption is problematic since several mechanisms can meddle in the formation of a functional gene product. Many TFs for example are embedded in transcriptional feedback loops (Amit et al., 2007; Segal et al., 2003) which results in changes in their mRNA levels. However, at least in systems where the regulation occurs mainly on the transcriptional level such as in *Drosophila melanogaster* (Segal et al., 2008), RNAP occupancy is indeed highly predictive for levels of transcription. Similarly, it was also shown that even TF-binding signals around the transcription start site are highly predictive of gene expression levels (Cheng et al., 2012).

In summary, thermodynamic models are based on equilibrium statistical mechanics. And, even though this framework comes with rigid assumptions, thermodynamic models have found a large area of applicability in biology, especially in the understanding of transcriptional regulation. In the present study, I apply the

thermodynamic framework to the *cis*-regulatory region of the *CYP1A1* promoter. In combination with a simple signal transduction model of the complex formation of the *CYP1A1*-binding TFs by the AhR and the Wnt/β-catenin signaling pathways, I uncover the transcriptional regulation of said promoter. With the help of a large library of synthetic and natural promoter constructs, this model is able to predict the importance of the two signal transduction pathways for the transcriptional regulation of the *CYP1A1* promoter. Furthermore, predictions on a physiological scale, i.e., at the level of the liver lobules are possible and highlight the molecular underpinnings of the portocentral expression gradient of *CYP1A1* in the liver.

2 Materials and Methods

A large part of this chapter has already been published in close collaboration with Alexandra Löffler, Silvia Vetter, Luisa Kreft and Albert Braeuning (Schulthess et al., 2015). The listed co-authors contributed with the following experimental work: AL carried out transfections and luciferase assays. LK quantified hepatic zonation. AB constructed the plasmids and performed Western Blotting, Immunoprecipitation and animal experiments.

2.1 Biological material and methods

2.1.1 Plasmids

Generation of a pT81luc-based Firefly luciferase reporter plasmid containing an approximately 1 200 bp fragment of the human *CYP1A1* promoter has been described previously by Schreiber et al. (2006). This plasmid contains four functional AhR-binding DREs, termed *C*, *D*, *E*, and *F* (Kress et al., 1998) and one TCF/β-catenin binding site, termed *T* (Braeuning et al., 2011). Different mutant versions of the reporter were generated by site-directed mutagenesis of individual TFBS's using the QuikChange kit (Stratagene, La Jolla, CA, USA) as recently described by Braeuning et al. (2011). An overview of the mutations introduced is given in Figure 2.1. Synthetic promoter constructs were generated containing multiple copies of either

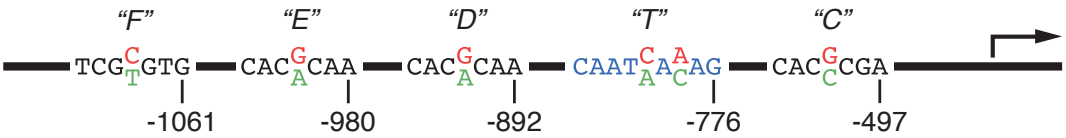


Figure 2.1: Structure of 5'-upstream regulatory region of the *CYP1A1* promoter. The binding site sequences of the DREs are shown in black and termed *C*, *D*, *E*, and *F*. The TCF7L2 binding site sequence is shown in blue and termed *T*. Single nucleotides were mutated (one mutation in the DREs and two in the TCF7L2 binding site) and are shown as changes from red to green. Furthermore, the positions relative to the transcription start site are shown.

C-DRE (sequence 5′-CGCTTCTCACGCGAGCCGG-3′) or D-DRE (sequence 5′-GCCGGCGCACGCAAGCTAG-3′) by cloning synthetic oligonucleotides into the SmaI site of pT81luc. Versions of the 2x C-DRE plasmid with different distances between the two DREs were generated by inserting non-AhR-responsive sectors of different size from the 1.2 kb *CYP1A1* promoter fragment between the two C-DREs of the NaeI-cut pT81-luc/2x C-DRE plasmid, resulting in 2x C-DRE variants with 49, 156, and 292 bp distance between the two AhR binding sites, respectively. The integrity of all plasmids was verified by dideoxy sequencing. In some experiments, a pCMV4-based expression vector for human AhR was used (Loeppen et al., 2005); control cells were transfected with empty pCMV4. Activity of the Wnt/β-catenin pathway was monitored with the SuperTOPFlash vector, which expresses Firefly luciferase under the control of 8x TCF/β-catenin binding sites (Braeuning et al., 2007b). Plasmid pRL-CMV encoding Renilla luciferase under the control of the constitutive cytomegaly virus promoter (Promega, Mannheim, Germany) was co-transfected and used for normalization of Firefly luciferase signals.

2.1.2 Cell culture and transfection

Mouse hepatoma cells from lines 55.1c (Braeuning et al., 2011), Hepa1c1c7 and the AhR-deficient sub-clone Hepa12 (Braeuning and Buchmann, 2009) were cultured in D-MEM/F-12 medium supplemented with 10 % fetal bovine serum and antibiotics (all reagents purchased from Invitrogen, Karlsruhe, Germany) at 37 °C and 5 % CO_2 in a humidified atmosphere. 55.1c cells carry a heterozygous deletion in exon 3 of *Ctnnb1*, encoding a constitutively active version of β-catenin. Cells were seeded on 24-well plates at a density of 40 000 cells/cm^2 24 h prior to transfection with Lipofectamine 2000 (Invitrogen). Cells were treated with the indicated concentrations of TCDD (Ökometric, Bayreuth, Germany) and/or the β-catenin inhibitors iCRT3 (Merck, Darmstadt, Germany), FH535 (Merck), or PNU74654 (Sigma, Taufkirchen, Germany) for 24 h, starting 24 h after transfection. All compounds were dissolved in dimethylsulfoxide and final concentration of the solvent in culture medium was 0.2 %. Cells were lysed with 1x Passive Lysis Buffer (Promega) for 15 min at room temperature. Dual-luciferase assays for Firefly and Renilla luciferase activities were conducted as recently described by Braeuning and Vetter (2012). Primary hepatocytes were obtained from young male adult mice with hepatocyte-specific knockout of *Ctnnb1* (encoding β-catenin) by standard collagenase perfusion (Braeuning

and Buchmann, 2009) and seeded at a density of 50 000 cells/cm^2 on 6-well plates coated with rat tail collagen in D-MEM/F-12 medium supplemented with 10 % fetal bovine serum and antibiotics. Medium was changed to 1 % serum after 6 h and cells were treated as described above.

2.1.3 Cell viability and growth analysis

All compounds were tested for the absence of cytotoxicity by the neutral red uptake and Alamar blue reduction assays as previously described by Braeuning et al. (2012). Cell growth was monitored by the use of the sulforhodamine B assay according to Skehan et al. (1990). All assays were conducted in octuple determinations on 96-well plates, where cells were seeded at 5 000 cells/cm^2 (sulforhodamine B assay) or 9 000 cells/cm^2 (cytotoxicity assays).

2.1.4 Immunoprecipitation and western blotting

Whole cell lysates were prepared according to Braeuning et al. (2011). Immunoprecipitation was performed at 4 °C over night using ProteinG-agarose beads and an antibody against AhR (Biomol, Hamburg Germany; 1:200 dilution). Whole cell lysates (50 µg/lane; protein concentration determined by use of the Bradford assay) or immunoprecipitates were separated by SDS-PAGE and transferred to PVDF membranes. Proteins were visualized by using antibodies against AhR (Biomol; 1:1 000), β-catenin (BD biosciences, Heidelberg, Germany; 1:500), or GAPDH (Merck; 1:1 000) in combination with appropriate alkaline phosphatase-conjugated secondary antibodies (Millipore, Schwalbach, Germany; 1:10 000) and the substrate CDP-Star (Tropix, Darmstadt, Germany). Chemoluminescence was monitored with a CSC camera (Raytest, Straubenhardt, Germany).

2.1.5 Gene expression analysis

Total RNA was isolated by TRIzol (Invitrogen) and reverse transcribed using avian myeloblastosis virus reverse transcriptase (Promega) as described previously by Braeuning and Vetter (2012). Real-time RT-PCRs were performed on a LightCycler system (Roche, Mannheim, Germany) using the FastStart DNA Master^{PLUS} SYBR Green I kit (Roche) according to the manufacturer's instructions and the following

primer pairs: 18s rRNA_fwd 5′-CGGCTACCACATCCAAGGAA-3′; 18s rRNA_rev 5′-GCTGGAATTACCGCGGCT-3′; Cyp1a1_fwd 5′-TGTCCTCCGTTACCTGCCTA-3′; Cyp1a1_rev 5′-GTGTCAAACCCAGCTCCAAA-3′; Cyp1a2_fwd 5′-GAGCGCTGTATCTACATAAACCA-3′; Cyp1a2_rev 5′-GGGTGAACATGATAGACACTATTGT-3′. Data was normalized according to the method described by Pfaffl (2001) with 18s rRNA as a housekeeping gene.

2.1.6 Animal experiment and immunostaining

Young adult male C3H/HeN mice (5 to 6 per group) were treated with the AhR inducer 3-MC (Sigma; dissolved in corn oil) by a single i.p. injection of 10, 25, or 50 mg/kg body weight 48 h prior to sacrifice. Mice were killed between 9 and 11 a.m. to avoid circadian variations; livers were excised, transferred to Carnoy's fixative and subsequently embedded in paraffin. Tissue slices of 5 µm thickness were stained for CYP1A as recently described by Braeuning et al. (2010) using a rabbit antiserum at 1:500 dilution (gift of Dr. R. Wolf, University of Dundee, UK) and a horseradish peroxidase-conjugated secondary antibody (1:100; Dako, Glostrup, Denmark) with the substrates 3-amino-9-ethylcarbazole/H_2O_2. Mice had access to tap water and standard chow ad libitum. All animals received humane care and protocols complied with institutional guidelines. Width of CYP1A-positive zones was assessed using an AxioImager light microscope and AxioVision software (Zeiss, Oberkochen, Germany).

2.2 Data processing and statistical analyses

2.2.1 Normalization of the basal reporter activity

The measured relative luciferase activity revealed significant differences in the basal levels, i.e., unstimulated or solvent controlled data points of both the natural and the synthetic reporter constructs (Figure 2.2). Normalization enables the comparability of effects of various stimulations and inhibitions on the reporter constructs. Thus, the measurements of each single-stimulated reporter construct were normalized individually to the unstimulated observations. The measurements of the reporter constructs stimulated with TCDD and a β-catenin inhibitor were normalized to 0 nM TCDD and 100 % β-catenin activity as this is the wild type state of the cells.

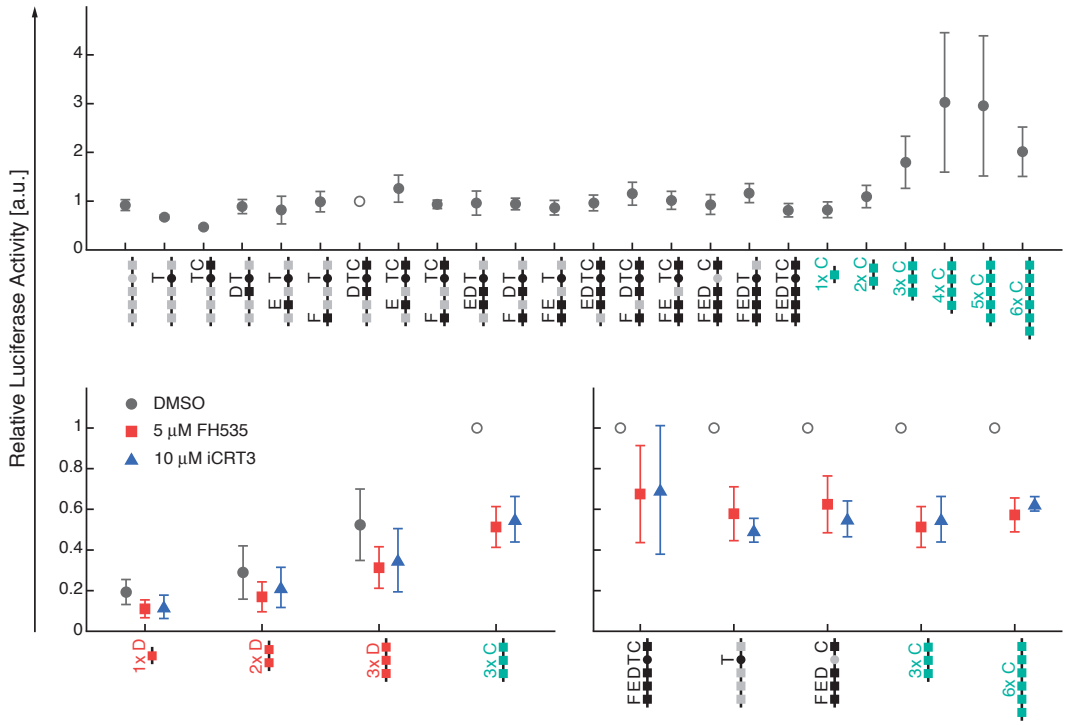


Figure 2.2: Basal reporter activity for natural and synthetic constructs before normalization. The hollow circles represent the reference data point to which the relative luciferase measurements were normalized. Gray data points represent untreated measurements while red and blue data points depict constructs treated with 5 µM or 10 µM of the β-catenin inhibitors FH535 and iCRTR3, respectively. Error bars represent one standard deviation of 4–10 biological replicates.

2.2.2 Correction of the standard deviations

Measurements of experimental data always are defective and at least in part inaccurate. To account for random errors and inconsistencies in the relative luciferase measurements the raw data were preprocessed as described in the following.

Measurement error model

A consequence of defective measurements is that the estimated means $\hat{\mu}$ and standard deviations $\hat{\sigma}$ from multiple replicates are defective as well. With errors-in-variables models also known as measurement error models such inaccuracies can be accounted for if the variables are independent. Since the cells holding the different constructs were stimulated separately the independence of the variables, i.e.,

reporter constructs is safe to assume. Therefore, I assumed that a first order relationship between the means and the standard deviations of every reporter construct c exist such that

$$\sigma_c(\mu_c) = a\mu_c + b \; . \tag{2.1}$$

With the help of a least squares approach the intercept a and the slope b were estimated. The corrected standard deviations σ^* corresponded to the arithmetic mean of the estimated standard deviations $\hat{\sigma}_c$ and the correct standard deviations σ_c. In the case of double-stimulated data, the estimated means and corresponding standard deviations of all stimuli combinations were merged. The corrected standard deviations were then jointly estimated for all stimulus combinations according to the method outlined above.

Lower threshold for the standard deviations

Because the standard deviations of the data were small for low concentrations and large for high concentration the parameter estimation algorithm was weighting low concentration data points too much while weighting high concentration data points too little. Thus, for a less stringent penalty to the fitting algorithm a lower threshold for the corrected standard deviations was set. Each standard deviation that was lower than 0.5 was set to 0.5. For the plots that just show the data without the fits of the mathematical model the standard deviations that were lower than 10 % of the corresponding mean were set to said value.

Standard deviation of the mean

Having already established the statistical independence of the samples, the standard deviation of the mean now informs about the precision of the means calculated from the replicates. With the help of

$$\sigma_c^{**} = \frac{\sigma_c^*}{\sqrt{n_c}} \; , \tag{2.2}$$

the newly corrected standard deviations σ_c^{**} for each reporter construct c and the previously corrected standard deviations σ_c^* can be calculated. n_c represents the number of biological replicates within the experiment.

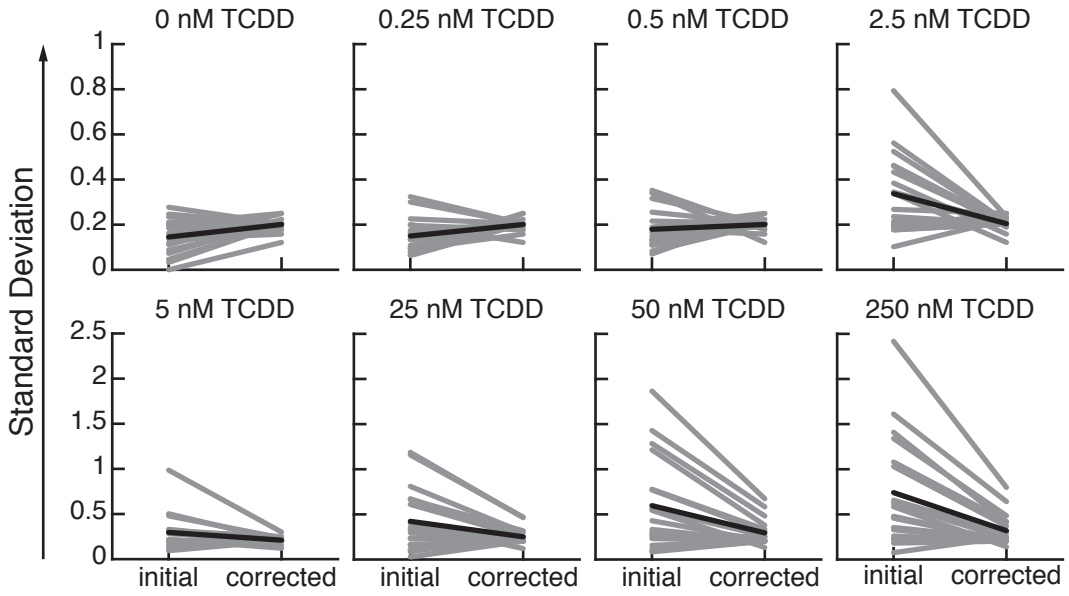


Figure 2.3: Change in standard deviations after the corrections. For each TCDD concentration (panels) the change in standard deviation before (initial) and after (corrected) the correction methods is shown. Gray lines represent the standard deviations of the 18 natural promoter constructs. Black lines depict the mean.

In conclusion, the correction of the standard deviation was necessary since random errors and inconsistencies in the experimental data were expected. In Figure 2.3 the overall effect of the aforementioned corrections is summarized. For each TCDD concentration and each of the 18 natural promoter constructs I plotted the change in standard deviation after the corrections. Furthermore, I calculated the mean before and after the correction procedure for all constructs of a specific TCDD concentration. While small TCDD concentration lead to a mean increase of the standard deviations, a mean decrease could be observed for high TCDD concentrations. In summary, random experimental errors and inconsistencies lead to smaller than expected standard deviations for small TCDD concentrations, and larger than expected standard deviations for high TCDD concentrations.

2.3 Bioinformatic analysis of TF binding site sequences

2.3.1 Representation of sequence alignments

A set of N aligned sequences can be represented by a so called position specific count matrix (PSCM) $\boldsymbol{C} = (C_{ij})$ in which the accumulated number i of the bases `A`, `C`, `G` and `T` for each position j is specified. The set of sequences

```
ATCAAAGC
ATCAAAGG
CTAAAGGG
TTCAAAGG
TTCAAAGT
TTCAAAGC
```
(2.3)

can thus be represented by

$$\boldsymbol{C} = \begin{array}{c} \\ \texttt{A} \\ \texttt{C} \\ \texttt{G} \\ \texttt{T} \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{bmatrix} 2 & 0 & 1 & 6 & 6 & 5 & 0 & 0 \\ 1 & 0 & 5 & 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 6 & 3 \\ 3 & 6 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array} . \tag{2.4}$$

The entry "5" in the second row and third column for example collects the `C`'s in each sequence at position 3. Note that incomplete sequencing snippets within the set of sequences can easily be accommodated with position specific numbers of observations N_j. Furthermore, inconclusive sequencing results such as `R`, meaning `A` or `G` within the IUPAC notation (IUPAC, 1971) can be represented by half counts in the PSCM.

Small samples often lead to zeros in the PSCM (cf. Equation 2.4). To correct for small sample sizes pseudocounts R are added. The corrected PSCM $\boldsymbol{C'}$ is therefore given by $\boldsymbol{C'} = \boldsymbol{C} + R$. Since they were introduced in the past (Durbin et al., 1998), no consensus on the optimal pseudocount value emerged. While a pseudocount of 1 is most commonly used, values differ from 0.01 (Chen et al., 1995) to up to 4 (Berg and Hippel, 1987). A recent study performing large scale simulations found the supposedly optimal pseudocount value according to various measures

and suggested the use of 0.8 (Nishida et al., 2009). But using a fixed value for the pseudocounts disregards the nucleotide distribution at a specific position and overestimates the occurrence of rare bases while underestimating more common bases. Therefore Rahmann et al. (2003) developed a correction method that is based on the ideas that (i) overall nucleotide composition of the motif should remain unchanged, and (ii) each position is regularized, i.e., corrected depending on its signal strength which leaves core motifs relatively untouched.

Rahmann regularization

At a fixed position j the letters are distributed according to

$$\tau_i \equiv \frac{C_{ij}}{N_j} \,. \tag{2.5}$$

The overall symbol distribution within the PSCM is therefore given by

$$\rho_i \equiv \frac{\sum_j C_{ij}}{\sum_j N_j} \,. \tag{2.6}$$

In order to keep the significant signal in τ intact when moving it towards ρ a deviation measure is defined as

$$\Delta(\omega) \equiv 2N_j \sum_i \delta(\omega)_i \ln\left(\frac{\delta(\omega)_i}{\rho_i}\right) \,. \tag{2.7}$$

Herein $\delta(\omega) \equiv (1-\omega)\tau + \omega\rho$ for $0 \leq \omega \leq 1$ is the family of regularized distributions and ω represents a suitable weight with which position j is regularized. The deviation measure Δ is a generalized log-likelihood statistic and its expectation E equals the number of its degrees of freedom. For large sample sizes and the alphabet of DNA (i.e. the four nucleotides `A`, `C`, `G` and `T`) it was shown that $E \to 3$. For small sample sizes the expectation is

$$E = 2\sum_i \sum_{k=0}^{N_j} \binom{N_j}{k} \rho_i^k (1-\rho_i)^{N_j-k} k \ln\left(k/(N_j\rho_i)\right) \,. \tag{2.8}$$

If the deviation measure for the unregularized distribution, i.e., for $\omega = 0$, is larger or equal the expectancy, an $\omega \in [0,1]$ needs to be found for which $\Delta(w) = \Delta(0) - E$.

If $\Delta(0) < E$, ω is set to 1 which means that the PSCM is corrected with its own overall symbol distribution. If a suitable regularizing weight ω was found the columns of the regularized PSCM $\boldsymbol{C'}$ follow to

$$C'_j = N_j \cdot \tau + \frac{\omega \cdot N_j}{1-\omega} \cdot \rho \ . \tag{2.9}$$

The regularization of the PSCM $\boldsymbol{C}$ in Equation 2.4 therefore is

$$\boldsymbol{C'} = \begin{array}{c} \\ \texttt{A} \\ \texttt{C} \\ \texttt{G} \\ \texttt{T} \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{bmatrix} 2.57 & 0.08 & 1.14 & 6.14 & 6.14 & 5.27 & 0.08 & 0.23 \\ 1.23 & 0.03 & 5.05 & 0.06 & 0.06 & 0.11 & 0.03 & 2.09 \\ 0.28 & 0.04 & 0.07 & 0.07 & 0.07 & 1.14 & 6.04 & 3.11 \\ 3.28 & 6.04 & 0.07 & 0.07 & 0.07 & 0.14 & 0.04 & 1.11 \end{bmatrix} \end{array} . \tag{2.10}$$

The probabilistic occurrences of the letters in the sequences is given by the profile $\boldsymbol{P} = (P_{ij}) = (C_{ij})/N_j$. For $\boldsymbol{C'}$ the profile follows to

$$\boldsymbol{P'} = \begin{array}{c} \\ \texttt{A} \\ \texttt{C} \\ \texttt{G} \\ \texttt{T} \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{bmatrix} 0.35 & 0.01 & 0.18 & 0.97 & 0.97 & 0.79 & 0.01 & 0.03 \\ 0.17 & 0.01 & 0.80 & 0.01 & 0.01 & 0.02 & 0.01 & 0.32 \\ 0.04 & 0.01 & 0.01 & 0.01 & 0.01 & 0.17 & 0.97 & 0.48 \\ 0.45 & 0.97 & 0.01 & 0.01 & 0.01 & 0.02 & 0.01 & 0.17 \end{bmatrix} \end{array} . \tag{2.11}$$

Here it should be noted that the profile assumes statistical independence between the positions as the probabilities of the positions are calculated independently from the remaining positions.

The profile now aids in the identification of other motifs with similar signals within a specific sequence. If the profile for example represents a binding sequence of a TF it can be used to find other binding sites for said TF within the genome. To make a meaningful decision whether a sequence window contains a signal or whether it is just background, a probabilistic model for the background is needed. The most simple background assumes an uniform distribution of the letters. Alternatively, the GC-content is often used to infer the background distribution. The decision if a signal is present is based on a likelihood ratio of the models, i.e., the ratio of the probabilities of the observed sequence in both models. For a sequence S

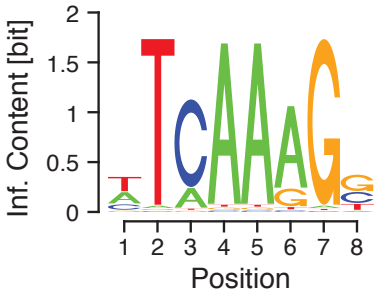


Figure 2.4: Sequence logo of $\boldsymbol{W}'$. The relative size of the letters represents their frequency in the sequences. The total height of the letters depicts the information at the position in bits.

of length L and a background distribution β, the profile $\boldsymbol{P}$ therefore translates into a position specific scoring matrix, more commonly known as a PWM $\boldsymbol{W} = (W_{ij})$ such that

$$W_{ij} \equiv \log_2 \left(\frac{P_{ij}}{\beta_i} \right) . \tag{2.12}$$

The score, i.e., the log odds score is then given by summing the scores within $\boldsymbol{W}$ representing the letters in S such that

$$\text{Score}(S) \equiv \sum_{i=1}^{L} W_{S_j,j} . \tag{2.13}$$

For the profile in Equation 2.11 and a uniform background distribution the PWM is

$$\boldsymbol{W}' = \begin{array}{c} \\ \texttt{A} \\ \texttt{C} \\ \texttt{G} \\ \texttt{T} \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{bmatrix} 0.5 & -4.2 & -0.5 & 2.0 & 2.0 & 1.7 & -4.2 & -2.9 \\ -0.6 & -5.6 & 1.7 & -4.8 & -4.8 & -3.9 & -5.6 & 0.4 \\ -2.7 & -5.2 & -4.5 & -4.5 & -4.5 & -0.6 & 2.0 & 0.9 \\ 0.8 & 2.0 & -4.5 & -4.5 & -4.5 & -3.6 & -5.2 & -0.6 \end{bmatrix} \end{array} . \tag{2.14}$$

As a graphical representation of the scores, sequence logos were introduced (Schneider and Stephens, 1990). Therein, the total height of the stacked letters depicts the information measured in bits at the corresponding position while the heights of the single letters represents their frequency in the set of sequences. The sequence logo of $\boldsymbol{W}'$ is shown in Figure 2.4. The consensus sequence for the set of sequences in Equation 2.3 is `WTCAAAGS`. For the sequence $S' =$ `ATCAAAGC` the log odds score is 12.3. Since the log odds score is much greater than zero the tested sequence contains a signal similar to the profile $\boldsymbol{P}'$.

2.3.2 Transcription factor binding affinity

Based on a biophysical model Foat et al. (2006) developed a method to predict the association constants between TFs and DNA. The resulting binding affinities are not quantitative but instead quantify the ratio of association constants between two DNA sequences that differ slightly.

Let S_{ref} be a reference DNA sequence for which I assume that it has the highest affinity to a TF. A DNA sequence that differs from S_{ref} by a point-mutation to base i at position j is denoted by S_{mut}. The position specific affinity matrix $\boldsymbol{A} = (A_{ij})$ is defined as

$$A_{ij} \equiv \frac{K_a(S_{\text{mut}})}{K_a(S_{\text{ref}})} = e^{-\frac{\Delta\Delta G_{ij}}{RT}} , \tag{2.15}$$

where $K_a(S)$ is the association constant, i.e., the binding affinity between the TF and the DNA sequence S. The change in Gibbs free energy in response to the point-mutation is defined as $\Delta\Delta G_{ij} \equiv \Delta G(S_{\text{mut}}) - \Delta G(S_{\text{ref}})$. For any sequence S of length L

$$K_a(S) = K_a(S_{\text{ref}}) \prod_{j=1}^{L} A_{S_j,j} \tag{2.16}$$

holds, if the change of binding free energy is additive, i.e., multiple point-mutations to the same sequence affect the affinity independently. If the concentration of the TF is furthermore far below the dissociation constant of the sequence with highest affinity and thus, selection is proportional to binding affinity such that $P_{ij} \propto A_{ij}$ it follows that

$$A_{ij} = \frac{P_{ij}}{P_{\text{max},j}} . \tag{2.17}$$

The most frequent base at position j is denoted by $P_{\text{max},j}$. In summary, it is possible to directly transfer a TF binding profile into a binding affinity.

For the regularized profile $\boldsymbol{P'}$ in Equation 2.11, the affinity matrix

$$\boldsymbol{A'} = \begin{array}{c} \\ \texttt{A} \\ \texttt{C} \\ \texttt{G} \\ \texttt{T} \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{bmatrix} 0.78 & 0.01 & 0.22 & 1.00 & 1.00 & 1.00 & 0.01 & 0.07 \\ 0.37 & 0.01 & 1.00 & 0.01 & 0.01 & 0.02 & 0.01 & 0.67 \\ 0.09 & 0.01 & 0.01 & 0.01 & 0.01 & 0.22 & 1.00 & 1.00 \\ 1.00 & 1.00 & 0.01 & 0.01 & 0.01 & 0.03 & 0.01 & 0.36 \end{bmatrix} \end{array} . \tag{2.18}$$

can be obtained. For the aforementioned sequence $S' = \texttt{ATCAAAGC}$ the change in

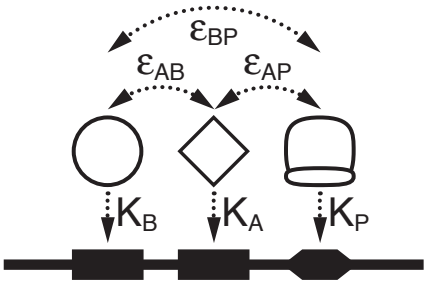


Figure 2.5: A thermodynamic model describing the binding of two activating TFs A and B to their respective binding sites as well as the association of a RNAP with DNA. K_i represent the association constants between i and the DNA, $\epsilon_i P$ describes the adhesive binding energy between TF i and the RNAP, and ϵ_{ij} represents the cooperative binding between the TFs i and j.

binding affinity calculates to 0.52. Thus, a TF is 48 % less likely to bind S' than the reference sequence $S_{\text{ref}} = \texttt{TTCAAAGG}$.

In Chapter 3.2, the here outlined method will be used to determine the log odds scores of the five TFBS's found in the *CYP1A1* promoter. Furthermore, I will calculate the reduction in binding affinity following the introduction of point-mutations within the core binding sequences of these TFBS's.

2.4 Matrix-based thermodynamic modeling

Thermodynamic models have long been used to describe the cooperative interactions of TFs in the *cis*-regulatory regions of genes (Ackers et al., 1982; Shea and Ackers, 1985). In the following I shortly outline a matrix-based formalism developed by Sherman and Cohen (2012) which I used for the implementation of the thermodynamic models of the *CYP1A1* promoter.

As introduced in Chapter 1.2.3, thermodynamic models relate binding site occupancies to expression. For a promoter that can be bound by two activators A and B as well as a RNAP this results in eight distinct microstates (Figure 2.5). Within the thermodynamic modeling framework these states are described by the binding probability p_{bound}. If only those microstate in which the RNAP is bound lead to transcription, the binding probability relates to the promoter occupancy, such that

$$p_{\text{bound}} = \frac{\text{transcribing states}}{\text{all possible states}} \,. \tag{2.19}$$

Sherman and Cohen (2012) introduced three descriptive matrices to decode the list of states. The position matrix $\boldsymbol{L}$ represents the microstates (rows) and holds the concentrations of the TFs and the RNAP associating with their respective binding sites (columns). For the example introduced in Figure 2.5 the position matrix is

$$\boldsymbol{L} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & [P] \\ 1 & [A] & 1 \\ 1 & [A] & [P] \\ [B] & 1 & 1 \\ [B] & 1 & [P] \\ [B] & [A] & 1 \\ [B] & [A] & [P] \end{bmatrix} . \tag{2.20}$$

The state vector $\boldsymbol{s}$ is the row-sum of $\boldsymbol{L}$ and the transcription vector $\boldsymbol{t}$ is a boolean vector determining which state is a transcribing state:

$$\boldsymbol{s}^T = \begin{bmatrix} 1 & [P] & [A] & [A][P] & [B] & [B][P] & [B][A] & [B][A][P] \end{bmatrix} \tag{2.21}$$

$$\boldsymbol{t}^T = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} . \tag{2.22}$$

With these three matrices all possible transcribing and not transcribing states of the promoter-TF-RNAP ensemble can be captured. Each state is subsequently weighted according to the Boltzman law. The statistical weight of a state is

$$\zeta([i],[j]) = \exp(-\epsilon_{ij}/k_B T) , \tag{2.23}$$

when ϵ_{ij} is its energy and k_B and T are the Boltzman constant and the absolute temperature, respectively. Note that ζ is just the concentration based reformulation of the statistical weight as defined in Chapter 1.2.3. Interaction of two simultaneously bound TFs is characterized by a cooperative binding free energy (e.g. ϵ_{AB}). Bound TFs furthermore interact with the bound RNAP through another adhesive binding free energy (e.g. ϵ_{PA} and ϵ_{PB}). Binding to the DNA by TFs and the RNAP is characterized through an association constant (e.g. K_P, K_A and K_B). Within the formalism of Sherman and Cohen (2012) the equilibrium constants for each state are combined

into the equilibrium constants vector $\boldsymbol{b}$ such that

$$\boldsymbol{b} = \begin{bmatrix} 1 \\ K_P \\ K_A \\ K_P\,K_A\,\zeta([P],[A]) \\ K_B \\ K_P\,K_B\,\zeta([P],[B]) \\ K_A\,K_B\,\zeta([A],[B]) \\ K_P\,K_A\,K_B\,\zeta([P],[A])\,\zeta([P],[B])\,\zeta([A],[B]) \end{bmatrix} \equiv \begin{bmatrix} \\ \vdots \\ \\ b_{PA} \\ \vdots \\ b_{PB} \\ b_{AB} \\ b_{PAB} \end{bmatrix} . \tag{2.24}$$

The binding probability is now

$$p_{\text{bound}} = \frac{\boldsymbol{s}^T\ (\boldsymbol{b} \times \boldsymbol{t})}{\boldsymbol{s}^T\,\boldsymbol{b}} , \tag{2.25}$$

where $\boldsymbol{b} \times \boldsymbol{t}$ is the pairwise element product. For the introduced example the binding probability is

$$p_{\text{bound}}([P],[A],[B]) = \frac{q_P + q_{PA} + q_{PB} + q_{PAB}}{1 + q_P + q_A + q_B + q_{PA} + q_{PB} + q_{AB} + q_{PAB}} . \tag{2.26}$$

with $q_i = K_i\,[i]$, $q_{ij} = b_{ij}\,[i][j]$ and $q_{ijk} = b_{ijk}\,[i][j][k]$. Different binding site layouts or binding schemes (e.g. the RNAP only associates with the DNA when both TFs are already bound to their TFBS's) require different matrices.

With the help of this matrix approach for the thermodynamic modeling framework, even complex promoter architectures are quickly to realize.

2.5 Profile likelihood estimation

The usage of mathematical models is key to the field of systems biology. In combination with experimental data they aim to optimally describe biological processes to predict testable or even untestable behaviors. Therefore, it is essential to determine how well the mathematical model can describe the data. The identifiability analysis developed by Raue et al. (2009) is able to detect functionally related parameters within the model and determine the quality of the experimental data. In addition, the profile likelihood estimation determines parameter confidences, i.e.,

how reliable the parameters were estimated from the model fitting routine. These methods can then be used to plan further experiments or reduce the complexity of the model (Kreutz et al., 2012).

2.5.1 Parameter estimation

A reaction network can be described by a set of ordinary differential equations

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{p}) \tag{2.27}$$

$$\boldsymbol{y}(t, p) = \boldsymbol{g}(\boldsymbol{x}(t), \boldsymbol{q}) + \boldsymbol{\epsilon}(t) \tag{2.28}$$

with the internal model states $\boldsymbol{x}$ describing the dynamics of n species, an external stimulus $\boldsymbol{u}$, a set of dynamic parameters $\boldsymbol{p}$, a m-dimensional mapping $\boldsymbol{g}$ of the internal model states to the observables $\boldsymbol{y}$ containing scaling parameters $\boldsymbol{q}$ and a normally distributed measurement noise $\boldsymbol{\epsilon}$. To fully quantify the model the set of parameters

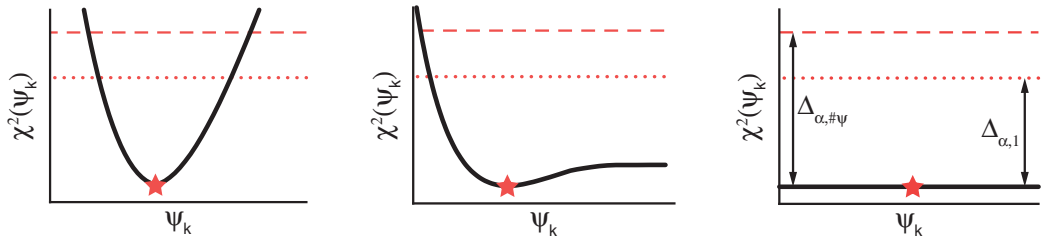
$$\boldsymbol{\psi} = \{\boldsymbol{p}, \boldsymbol{x}(0), \boldsymbol{q}\} \tag{2.29}$$

is needed wherein $\boldsymbol{x}(0)$ specifies the initial conditions of the model states. Note that for the model to be fully observable n has to be equal to m.

Usually the model parameters are not known and have to be estimated from experimental data. The agreement between the observables as predicted by the model and the experimental data $\boldsymbol{z}$ is commonly measured by the weighted sum of squared residuals

$$\chi^2(\boldsymbol{\psi}) = \sum_{i=1}^{m} \sum_{j=1}^{d} \left(\frac{z_{ij} - y_i(\boldsymbol{\psi}, t_j)}{\sigma_{ij}} \right)^2 , \tag{2.30}$$

where d is the number of data-points for each observable i and σ_{ij} represents the corresponding measurement error. $y_i(\boldsymbol{\psi}, t_j)$ is the i-th observable as predicted by the parameters $\boldsymbol{\psi}$ for time-point t_j. The parameters can now be estimated by finding the parameter values $\hat{\boldsymbol{\psi}}$ that minimizes $\chi^2(\boldsymbol{\psi})$. If the measurement noise is distributed normally, the weighted sum of squared residuals is proportional to the log-likelihood and minimizing $\chi^2(\boldsymbol{\psi})$ corresponds to the maximum likelihood estimation of $\boldsymbol{\psi}$ and $\chi^2(\boldsymbol{\psi}) = \text{const} - 2 \cdot \log(\mathcal{L}(\boldsymbol{\psi}))$ with $\mathcal{L}(\boldsymbol{\psi})$ representing the likelihood (Seber and Wild, 2003).



(a) ψ_k identifiable

(b) ψ_k practical non-identifiable

(c) ψ_k structural non-identifiable

Figure 2.6: Identifiability scenarios for the parameter ψ_k according to the profile likelihood $\chi^2(\psi_k)$ (solid lines). The point-wise and simultaneous thresholds ($\Delta_{\alpha,1}$ and $\Delta_{\alpha,\#\psi}$) are shown as dotted and dashed red lines, respectively. The optimal parameter is marked with a red star.

2.5.2 Confidence intervals

It is now assumed that an appropriate model which describes the existing experimental data is given. In order to make a statement about the uncertainty of the parameter estimates, i.e., their confidence intervals (Lehmann and Casella, 1998) the shape of the likelihood needs to be analyzed. A confidence interval $[\sigma_k^-, \sigma_k^+]$ of a parameter estimate $\hat{\psi}_k$ to a confidence level α signifies that the true parameter value ψ_k^* lies within this interval with probability α. For most biological applications finite sample confidence intervals are used. Here, a threshold $\Delta_{\alpha\lambda} = \chi^2(\alpha, \lambda)$ in the likelihood defines a confidence region $\{\boldsymbol{\psi} | \chi^2(\boldsymbol{\psi}) - \chi^2(\hat{\boldsymbol{\psi}}) < \Delta_{\alpha\lambda}\}$ whose borders represent confidence intervals (Meeker and Escobar, 1995). The choice of the degrees of freedom results in two different types of confidence intervals (Press et al., 1992). While setting λ to the number of parameters results in simultaneous confidence intervals that hold jointly for all parameters, $\lambda = 1$ yields point-wise confidence intervals that hold individually for each parameter.

2.5.3 Identifiability

If the confidence interval $[\sigma_k^-, \sigma_k^+]$ is finite, i.e., if $\chi^2(\psi_k)$ has a unique minimum at $\hat{\psi}_k$, the parameter ψ_k is called identifiable (Figure 2.6a). A non-identifiable parameter can be either practical or structural non-identifiable. Practical non-identifiability arises from limited amount or quality of experimental data used for the parameter estimation and results in an infinite confidence interval (Figure 2.6b). Therefore,

increasing the amount or quality of the experimental data can turn a practical non-identifiable parameter into an identifiable one. Structural non-identifiability on the other hand is independent of the experimental data but arises from the model structure alone, e.g. from a redundant parametrization (Cobelli and DiStefano, 1980). The confidence intervals of a structural non-identifiable parameter are infinite (Figure 2.6c). A structural non-identifiability can only be resolved if the model structure is altered or if new measurements, e.g. of an additional species are performed (Raue et al., 2011).

With the here outline method, I determined the identifiability of the parameters in the models for the synthetic as well as the natural promoter constructs (Chapters 3.4.1 and 3.5.1).

2.6 Methodological workflow and numerical framework

The bioinformatic analysis was done with the software RStudio Version 0.98.1103 which utilized R version 3.1.3. The databases used to obtain the PSCMs for the different binding sites were Transfac Rel. 12.1 and JASPAR 2014. For all remaining implementations and simulations I used MATLAB R2013a. For that the experimental data was imported from .xls files and pre-processed as described above. The intercept and the slope of the measurement error models were estimated with the polyfit algorithm. The combination of the equations of the signaling model and the construct specific binding probabilities served as objective functions for a weighted non-linear least square fit (trust-region-reflective algorithm in lsqnonlin). I utilized a Monte-Carlo algorithm in combination with latin hypercube sampling (McKay et al., 1979) to generate a set of 10 000 starting values for the optimization. The ranges of the parameters were limited to be biologically feasible (e.g. association constants greater than 0) and it was made sure that the resulting optima did not exceed any artificial boundaries such as an exceeding number of iterations. The identifiability and confidences of the parameter estimates was then assessed by exploiting the profile likelihood as described above. The so identified set of optimal parameters was then used to predict the integration of the two signaling pathways into the promoter. These predictions were in the end validated with experimental data. Figure 2.7 outlines the fitting strategy and highlights how the models for the synthetic and natural constructs are connected.

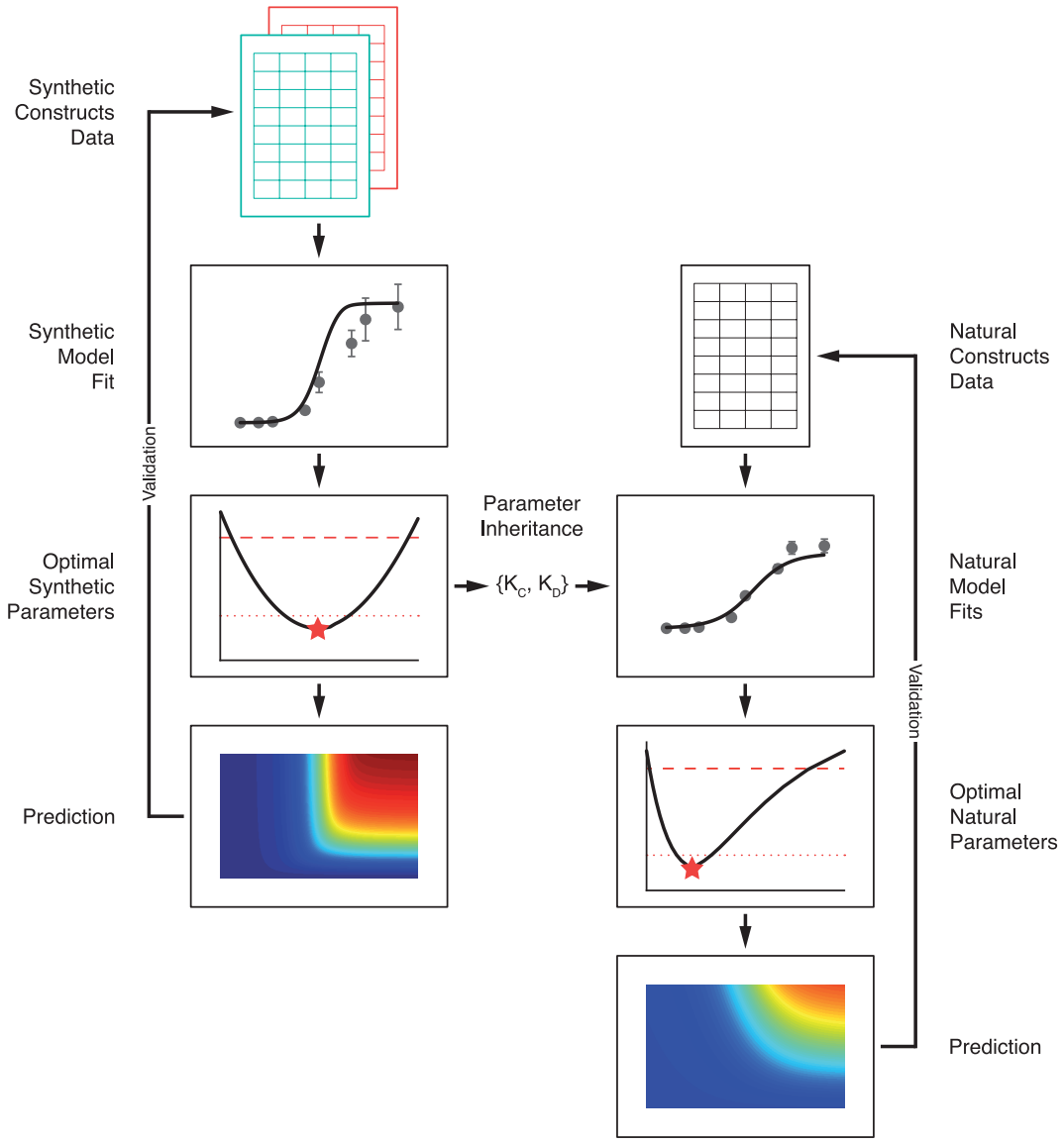


Figure 2.7: Methodological workflow. Following a parameter estimation and optimality analysis of the synthetic constructs the dual signal integration is predicted, and validated against the data. The parameter estimation of the natural constructs inherits the parameters K_C and K_D from the synthetic promoter model. Following the optimality analysis, the natural promoter is used to predict the dual signal integration. Subsequently, this prediction is validated against the natural construct data.

3 Results

A large part of this chapter has already been published in close collaboration with Alexandra Löffler, Silvia Vetter, Luisa Kreft and Albert Braeuning (Schulthess et al., 2015). The listed co-authors contributed with the following experimental work: AL carried out transfections and luciferase assays. LK quantified hepatic zonation. AB constructed the plasmids and performed Western Blotting, Immunoprecipitation and animal experiments.

Zonated gene expression following toxin exposure is a key feature of liver function. The underlying molecular mechanisms including the interplay of signal transduction pathways with promoter occupancy leading to hepatic zonation are still poorly understood. The present study exemplifies how mathematical modeling in connection with combinatorial reporter assays has the capacity to unravel these mechanisms.

To do so, I firstly disentangled how the AhR and Wnt/β-catenin signal transduction pathways affect the expression of the *CYP1A1* gene (Chapter 3.1). Based on the genomic architecture of the *CYP1A1* promoter and with the help of high throughput data and publicly available databases, I scored the TFBS's within the promoter and subsequently analyzed the impact of site-specific point-mutations on the binding affinity of TFs (Chapter 3.2). Following the experimental discovery of cooperativity between two TFs within the *CYP1A1* promoter (Chapter 3.3), I analyzed this cooperative behavior in more detail in Chapter 3.4 by combining data of synthetic promoter constructs with a comprehensive mathematical model. This model was then used to predict the integration of the two signal transduction pathways into the promoter. Building on the knowledge gained from the synthetic constructs I finally determined TF cooperativity and dual signal integration as well as their implications for hepatic zonation for the human *CYP1A1* promoter in Chapter 3.5.

3.1 AhR and Wnt/β-catenin signaling affect *CYP1A1* expression

Among the DREs in the promoter region of *CYP1A1* a binding site for TCF/β-catenin is located (Braeuning et al., 2011). Whether or not this TFBS is essential for Wnt-mediated expression of *CYP1A1* can be tested with a mutant reporter construct. Therefore, we inactivated this TCF/β-catenin binding site by point-mutation (cf. Figure 2.1) and transfected this reporter into cells of mouse hepatoma cell line 55.1c. Exposition to increasing concentrations of TCDD for 24 h and subsequent luciferase activity measurements resulted in a 3.5-fold induction (Figure 3.1).

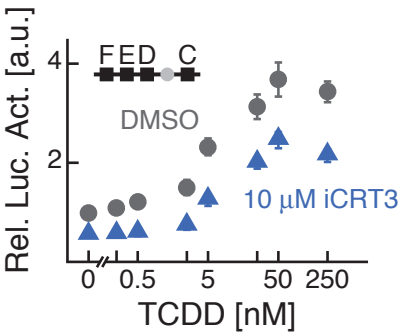
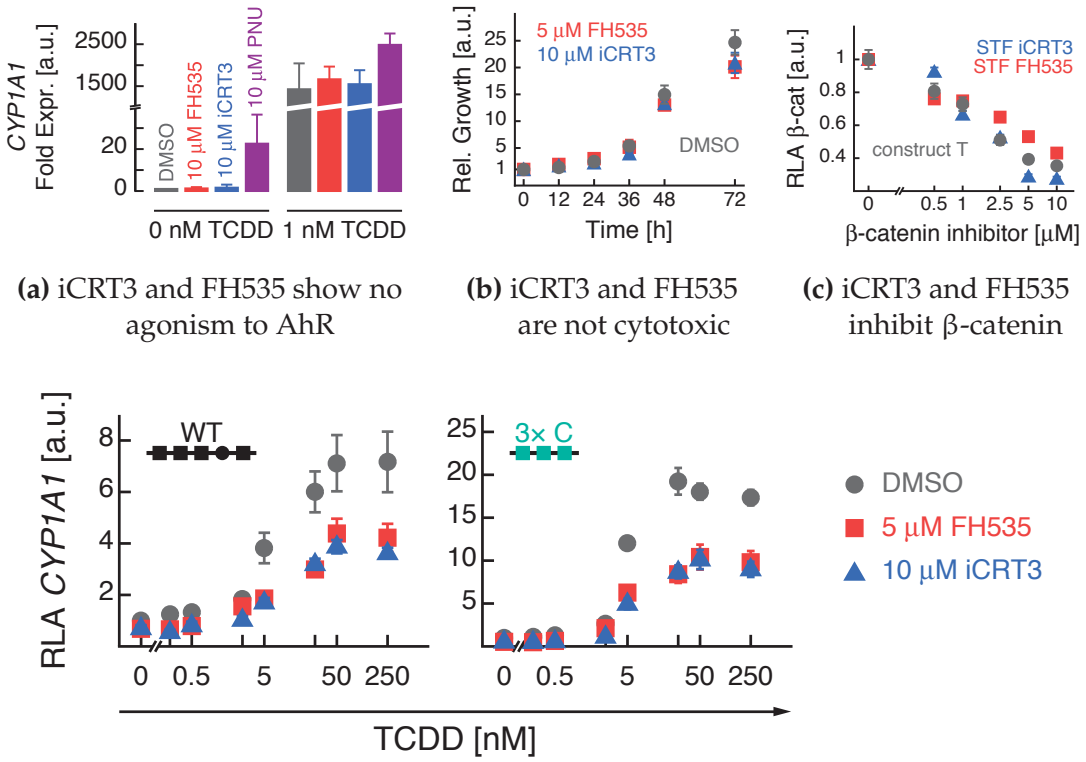


Figure 3.1: β-catenin modulated *CYP1A1* expression independently of the TCF/β-catenin binding site. Relative luciferase activity of a promoter construct in which the TCF/β-catenin binding site was inactivated by point-mutation over a series of TCDD concentrations treated with DMSO (gray) or 10 µM of the β-catenin inhibitor iCRT3 (blue).

Since 55.1c cells express a mutant, constitutively active form of β-catenin (Braeuning et al., 2011), β-catenin activity can be modulated by an inhibitor (Figure 3.2) To ensure that the inhibitor does not interfere with the expression of *CYP1A1* as mediated by the AhR signaling pathway, we tested three potential candidates (iCRT3, FH535 and PNU74654) for agonism to AhR in primary hepatocytes derived from β-catenin knockout mice (Braeuning et al., 2009). Independent of the presence or absence of TCDD, the two inhibitors iCRT3 and FH535 displayed no effect on *CYP1A1* expression as compared to a solvent control (Figure 3.2a). The application of PNU74654 on the other hand resulted in increased *CYP1A1* mRNA implying weak agonistic activity against AhR. The two non-AhR-agonistic inhibitors were subsequently tested for cytotoxicity by measuring the relative growth of 55.1c cells following treatment with FH535 or iCRT3. Reassuringly, both β-catenin inhibitors did not impede cell growth (Figure 3.2b). To calibrate the concentrations of the β-catenin inhibitors, we used the 8x TCF/β-catenin driven SuperTOPFlash reporter (cf. Figure 2.1) and measured the relative luciferase activity as a proxy for Wnt/β-catenin pathway activity (Figure 3.2c). It



(a) iCRT3 and FH535 show no agonism to AhR

(b) iCRT3 and FH535 are not cytotoxic

(c) iCRT3 and FH535 inhibit β-catenin

(d) iCRT3 and FH535 inhibit *CYP1A1* expression

Figure 3.2: Investigation of β-catenin inhibitors. **(a)** *CYP1A1* mRNA of hepatocytes from β-catenin knockout mice following treatment with DMSO (gray), 10 µM FH535 (red), 10 µM iCRT3 (blue) and 10 µM PNU74654 (purple) with or without 1 nM TCDD. **(b)** Relative growth of 55.1c cells following treatment with DMSO (gray circles), 5 µM FH535 (red squares) or 10 µM iCRT3 (blue triangles) over 72 hours. **(c)** Relative luciferase activity of a construct containing only a wild type TCF/β-catenin site (gray circles) and two SuperTOPFlash reporters of iCRT3 (blue triangles) and FH535 (red squares) after treatment with increasing concentrations of iCRT3 and FH535. **(d)** TCDD concentration series of WT and 3x C-DRE reporter constructs following treatment with DMSO (gray circles), 5 µM FH535 (red squares) or 10 µM iCRT3 (blue triangles).

can be seen that increasing concentrations of β-catenin inhibitor reduced the relative luciferase activity by up to 72 % for iCRT3. Furthermore, the induction of a reporter construct containing only a wild type TCF/β-catenin binding site was similarly reduced by 76 %. Thus, the inhibitors efficiently blocked Wnt/β-catenin signaling output. It should also be noted that since the activity of β-catenin was almost the same for 5 µM and 10 µM, I omitted the 5 µM data point from the following analyses. Furthermore, relative luciferase measurements for increasing TCDD concentrations showed an efficient inhibition of *CYP1A1* induction for both

the natural wild type and a synthetic reporter construct (Figure 3.2d). Since the inhibitory performance of iCRT3 was slightly better than the one of FH535 (72 % vs. 67 % inhibition in Figure 3.2c), we chose iCRT3 to inhibit β-catenin activity for all subsequent experiments. Having established iCTR3 as an effective modulator of the Wnt/β-catenin pathway, we could then apply this inhibitor to the reporter construct without a TCF/β-catenin TFBS of Figure 3.1 and measure the relative luciferase activity. 10 μM iCRT3 yielded an halved basal reporter activity as well as a reduction of *CYP1A1* induction (3.5-fold to 2-fold). Thus, a β-catenin inhibitor could even reduced the *CYP1A1* induction of a construct with an inactive TCF/β-catenin binding site, hinting at an additional mechanism of interaction between the AhR and the Wnt/β-catenin pathways.

3.1.1 Physical interaction between AhR and β-catenin

Recently, it was reported that there exists a direct physical interaction between AhR and β-catenin (Braeuning et al., 2011; Kasai et al., 2013). We could confirm this interaction with an co-immunoprecipitation (Figure 3.3). A western analysis of AhR expression in lysates of 55.1c cells resulted in equal bands for the solvent control, a stimulation with 250 nM TCDD and a stimulation with 250 nM TCDD and 10 μM of iCRT3. Thus, neither TCDD nor iCRT3 significantly affected AhR concentration. A subsequent western analysis of wild type β-catenin and a constitutively active β-catenin mutant after immunoprecipitation with an anti-AhR antibody under aforementioned stimulations was performed. We observed the strongest interaction between AhR and β-catenin when TCDD was applied and a reduction when β-catenin activity was inhibited with iCRT3. This observation could be made for both, the wild type and the exon 3-deleted, i.e., the constitutively active version of β-catenin. Since β-catenin is constantly degraded when its pathway is not activated by Wnt ligands (Niehrs, 2012), the bands for the wild type version of β-catenin were far less pronounced than those of the constitutively active β-catenin. Thus, the interaction between AhR and β-catenin can be modulated with both TCDD and iCRT3. The presented evidence of physical interaction between AhR and β-catenin indicated that the TF complex associating with the DREs in the *CYP1A1* promoter not only consists of TCDD, AhR and Arnt but also of β-catenin. This enables the AhR and the Wnt/β-catenin pathways to regulate *CYP1A1* expression through the amount of AhR and/or β-catenin available for TF complex formation. Since we

were using 55.1c cell which harbor constitutively active β-catenin, the only way by which the amount β-catenin could be modulated was through specific inhibitors such as iCRT3 (cf. Figure 3.1).

Now the question remained if there is enough AhR in the cells such that it is not limiting the production of TFs and therefore *CYP1A1* expression. By performing AhR overexpression experiments we observed that alterations in AhR expression did not remarkably influence DRE-driven reporter activities (Figure 3.4). For these experiments we utilized an empty as well as an AhR expression vector and co-transfected them with various reporter constructs under different conditions. The functionality of the expression vector was assessed with a western analysis which showed that the overexpression of AhR with its expression vector indeed resulted in higher AhR expression (Figure 3.4a). Hence, we co-transfected wild type and synthetic 3x C-DRE reporter constructs with both expression vectors and treated them with DMSO, 5 μM FH535 and 10 μM iCRT3 in the presence and absence of TCDD (Figure 3.4b). We observed only slight changes in reporter activity. The same holds true for a co-transfection with six synthetic reporter constructs either unstimulated or treated with 250 nM TCDD (Figure 3.4c). As a positive control we repeated the experiments in the AhR-deficient Hepa1c1 clone c12 which reassuringly resulted in elevated reporter activity for the AhR expression vector. From these experiments we concluded that the amount of AhR was indeed not rate limiting for the regulation of *CYP1A1*, and that the modulation of AhR mRNA expression by β-catenin was not relevant for *CYP1A1* regulation under our experimental conditions.

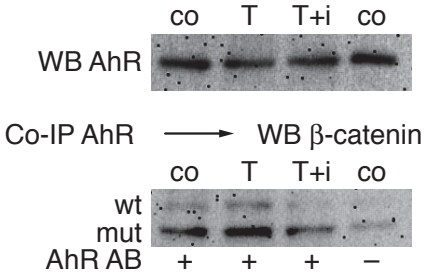
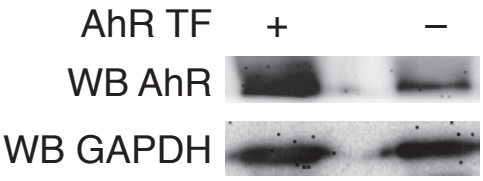
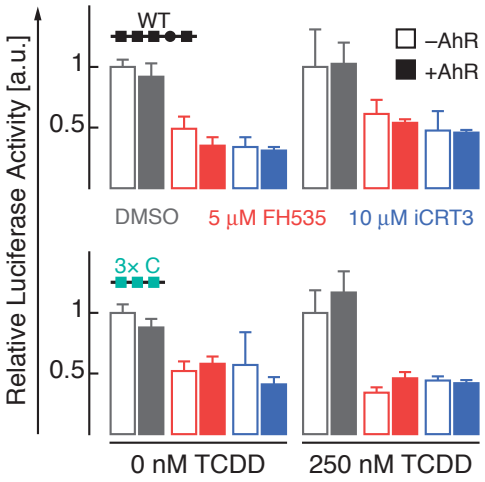


Figure 3.3: The interaction between AhR and β-catenin is increased upon TCDD treatment. Top: Western blot of AhR expression in lysates of 55.1c cells. Co: solvent control; T: 250 nM TCDD for 1 h; T+i: 250 nM TCDD + 10 μM iCRT3. Bottom: Western blot of β-catenin after IP with anti-AhR antibody (AhR AB "+"). Wild type (wt) and exon 3-deleted (mut) versions of β-catenin are shown.

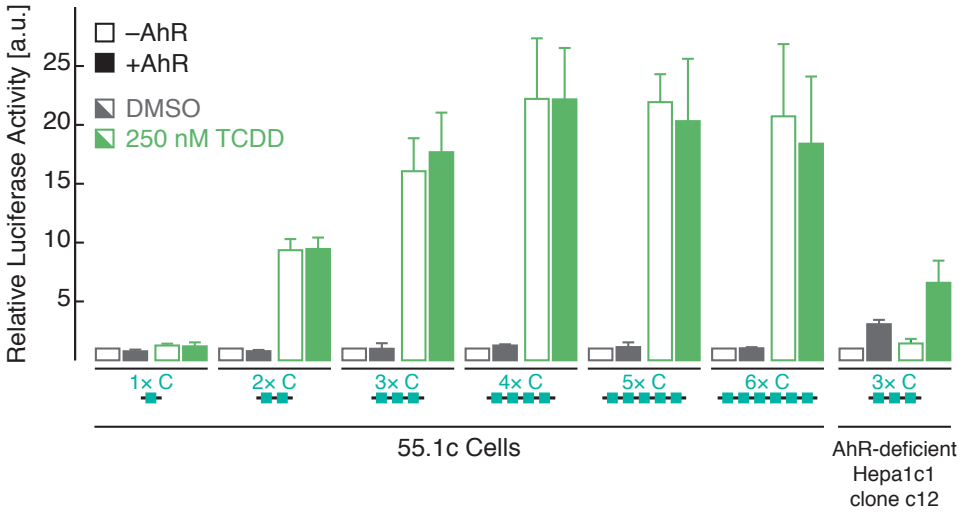
In summary, Wnt/β-catenin and AhR signaling are linked in two ways to regulate *CYP1A1*: (i) pathway specific TFs act on the same promoter and, (ii) AhR and β-catenin interact physically. An interesting aspect was that an interaction even existed without a functional TCF/β-catenin binding site. Thus, interactions



(a) Western analysis of AhR expression



(b) AhR co-transfection under β-catenin inhibitors



(c) AhR co-transfection of synthetic constructs

Figure 3.4: Alterations in AhR expression did not remarkably influence DRE-driven reporter activities. **(a)** Western analysis of AhR protein expression in 55.1c cells transfected with pCMV4/AhR (AhR TF "+") or empty pCMV4 ("-"). GAPDH was used as a loading control. **(b)** WT and 3x C-DRE reporter were co-transfected with either an empty (empty bars) or an AhR expression vector (filled bars). Treatment was with β-catenin inhibitors FH535 (red) and iCRT3 (blue) or DMSO (gray) and in the absence or presence of TCDD. Values obtained with the DMSO controls and the TCDD-treated cells were each normalized to 1 separately. **(c)** AhR co-transfection. 1x C-DRE to 6x C-DRE constructs either co-transfected with an empty (empty bars) or with an AhR expression vector (filled bars). Treatment was with 250 nM TCDD. Same setting in the AhR-deficient Hepa1c1 clone c12, the latter serving as a positive control for the functionality of the AhR expression vector.

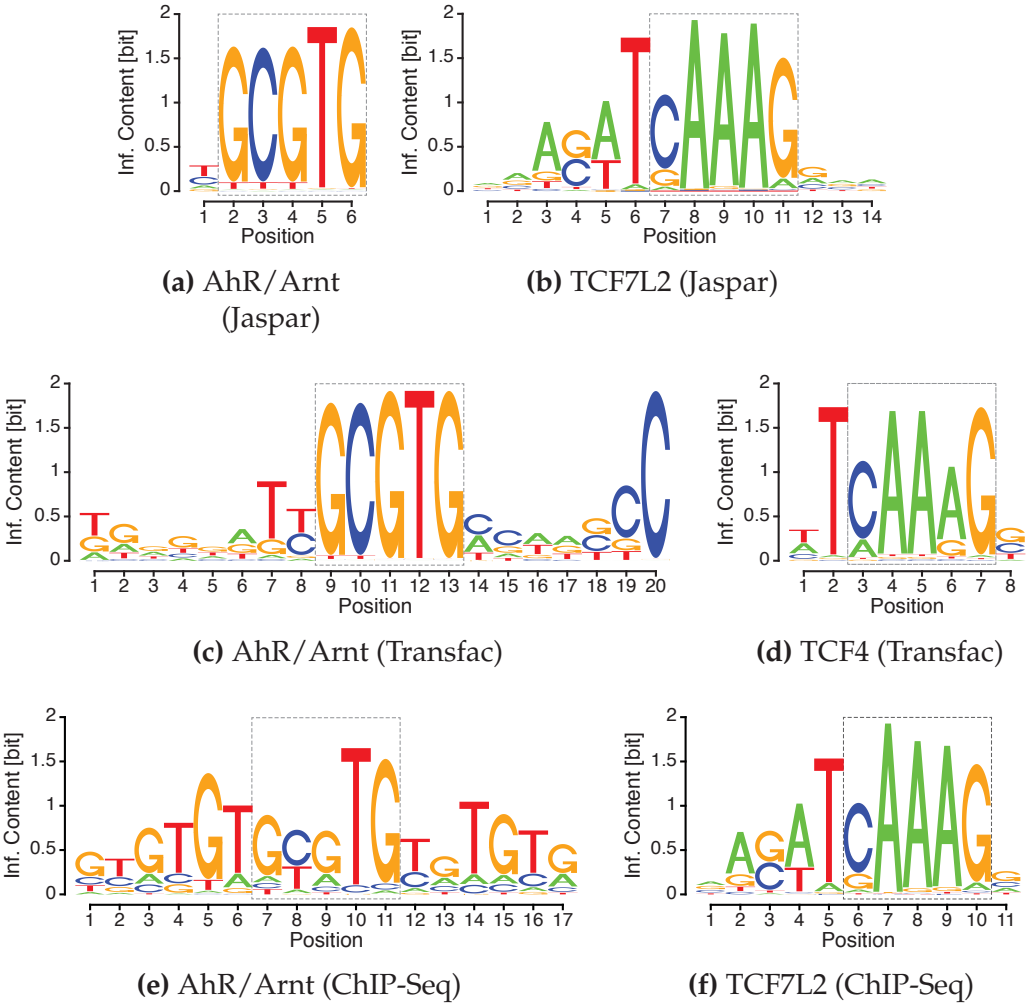
on the *CYP1A1* promoter seemed to be not necessary to establish a signaling level crosstalk, and it remained unclear which level of interaction would be the dominating mechanism by which these two pathways collaborate to regulate *CYP1A1* expression. To dissect this complex regulation of *CYP1A1* in more detail, the quantitative behavior of the response to AhR signaling as well as the quantitative interaction of both pathways will be studied. But, prior to that, I will dissect the architecture of the human *CYP1A1* to determine the effect of point-mutations on TF binding affinity with the help of a bioinformatic method making use of high-throughput data and publicly available databases.

3.2 Point-mutations influence TF binding affinity

Before analyzing the regulation of the *CYP1A1* promoter in more detail an understanding of the association of TFs with their target binding sites was necessary.

By binding to specific DNA sequences TFs can regulate a wide variety of cellular processes in particular the expression of genes. In combination with experimental high-throughput methods such as chromatin immunoprecipitation followed by sequencing (ChIP-Seq) *in silico* approaches have been developed to identify binding sites and sequences (Bussemaker et al., 2007; Stormo and Zhao, 2010). The binding motifs are subsequently derived from a set of aligned and functionally related sequences. Commonly, they are represented by their consensus sequence or a PWM (Stormo et al., 1982; Chapter 1.2.2). If the right conditions are met (e.g. binding site selection is proportional to binding affinity) it is even possible to calculate the energy with which a TF associates with a binding site.

In the following, I analyzed the five functional TFBS's within the *CYP1A1* promoter region in order to quantify the binding affinities of the corresponding TFs. The theoretical framework by which this was done was outlined in Chapter 2.3. Three different sources of sequence alignment data were used. Lo and Matthews (2012) performed ChIP-Seq experiments in breast cancer cells, mapped the binding sites of AhR and Arnt and kindly provided me with the sequence list of the AhR/Arnt cobound regions. In the context of the ENCODE project (ENCODE Project Consortium, 2012), Wang et al. (2012) determined the PWM of the Wnt/β-catenin target TCF7L2 which I used in the following analysis. Apart from these



(a) AhR/Arnt (Jaspar)

(b) TCF7L2 (Jaspar)

(c) AhR/Arnt (Transfac)

(d) TCF4 (Transfac)

(e) AhR/Arnt (ChIP-Seq)

(f) TCF7L2 (ChIP-Seq)

Figure 3.5: Sequence logos of the AhR/Arnt **(a, c, e)** and TCF/β-catenin **(b, d, f)** binding sites as derived from primary ChIP-Seq data and the databases Transfac and Jaspar. The total height of the stacked letters depicts the information content at the corresponding position measured in bit while the heights of the single letters represents their frequency in the set of sequences from which they were calculated. The dashed gray boxes enclose the core binding sequence of the corresponding TFs.

datasets, the TF binding profile databases Transfac (Matys et al., 2006) and Jaspar (Mathelier et al., 2014) also contain sequence alignments for AhR/Arnt and TCF7L2. It should be noted that the here presented binding site sequences of the C-, D- and E-DRE do not match the ones presented in the promoter structure earlier 1.4. This is the case because these binding sites are located on the reverse strand of the DNA. But, in order to uphold comparability between the DREs, I used the reverse complements of the aforementioned binding sites in the upcoming analysis. In the case of Transfac, I used a combination of the matrices M00235 and M00237 which

are both based on experiments in rabbits by Swanson et al. (1995). For TCF7L2 I used the matrix M00671. Note that the names TCF7L2 and transcription factor 4 (TCF4) were used synonymously. The Jaspar database provided PSCMs for murine AhR/Arnt and human TCF7L2. Where necessary, all data was translated into PSCMs and regularized according to the method of Rahmann et al. (2003).

For most of the data sets a distinct core binding region can be identified (Figure 3.5). Even though the AhR/Arnt logo from Jaspar is too short to draw conclusions about the flanking regions, it displayed a clear signal for the DRE core sequence 5′–GCGTG–3′ (Figure 3.5a). For the AhR/Arnt logo from Transfac this core region also exhibited a significantly higher information content as the flanking regions (Figure 3.5c) which lead to an extended DRE core sequence 5′–TNGCGTG–3′ (Swanson et al., 1995; Whitlock et al., 1996; Yao and Denison, 1992). In contrast, the most recent high-throughput experiments shown in Figure 3.5e also yielded an equal DRE core sequence (5′–GCGTG–3′) but a symmetrical extended DRE in the form 5′–GTGCGTG–3′ (Lo and Matthews, 2012). Additionally, only the ChIP-Seq data displayed the dimeric binding of AhR/Arnt that is visible by the two peaks in information content at positions 5 and 10 (Bacsi et al., 1995). The recognition sequence 5′–SATCAAAGS–3′ of the TCF7L2 TF (Cadigan and Waterman, 2012) was present in all three data sets (Figures 3.5b, 3.5d and 3.5f). Only small differences in information content and letter frequency could be observed for the core binding sequence as well as the flanking regions. Only the TCF4 logo from Transfac displayed a smaller and slightly dissimilar flanking region.

It is well known that the sequences flanking the core binding regions can affect binding affinity (Siggers and Gordân, 2014). Therefore, I utilized all the positions within the different motifs to calculate the log odds score against the binding sites in the 5′-upstream region of the *CYP1A1* gene (Kubota et al., 1991). Figure 3.6 shows the scores of the AhR/Arnt targeting the C- to F-DREs as well as those of TCF7L2 for the three different data sets. I additionally included the scores of the mutated and therefore supposedly inactivated binding sites that were part of our library of mutated promoter constructs (see Chapters 3.4 and 3.5). The first observation that could be made was that a mutation in the binding sequence always lead to a lower score. The short motifs (AhR/Arnt from Jaspar and TCF4 from Transfac) resulted in the highest scores. These motifs contained almost no flanking regions and high

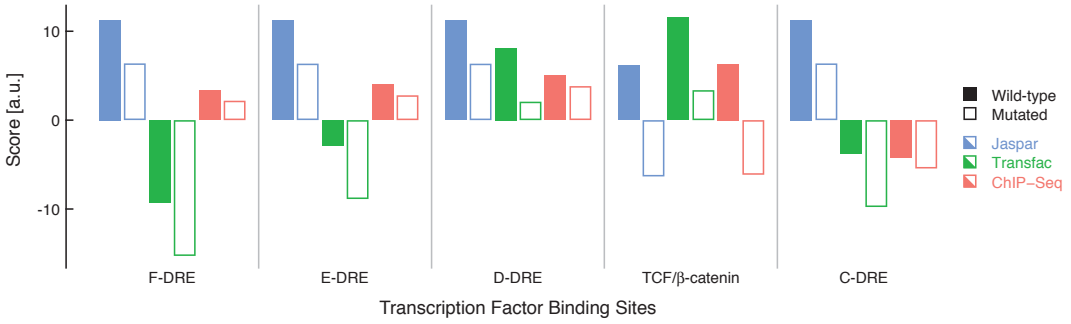


Figure 3.6: Scores of the various TFBS's of the *CYP1A1* promoter. The scores were calculated based on full length of the three source data sets (Jaspar, Transfac and ChIP-Seq) and for the wild type and the point-mutated binding site sequence as they were used in the experiments.

information contents in the core binding sites. This biased score was therefore expected. Surprisingly, the Transfac motifs scored against the DREs tended to lead to negative scores which may be due to the high frequency of C's at position 20. Similarly, the Transfac score of the D-DRE was only positive because there exists a cytosine at the position in the 5'-upstream region of the *CYP1A1* gene that corresponds to position 20 in the motif. The almost identical Jaspar and ChIP-Seq motifs of TCF7L2 also lead to very similar scores. And lastly, scoring the Transfac and ChIP-Seq motifs against the C-DRE sequence resulted in negative scores.

With the help of the position specific affinity matrix (cf. Chapter 2.3), the reduction of binding affinity in response to point-mutations in the core binding sequence could be calculated (Figure 3.7). Note that the core binding sequence was the same for all DREs (cf. Figure 2.1) Hence, I only calculated and displayed the binding affinity reduction for one DRE. While the binding energy was reduced almost 100 % for both groups of binding sites when using the PWMs from Transfac and Jaspar, the binding affinity towards DREs was only reduced by approximately 40 % when using the ChIP-Seq data.

Taken together this bioinformatic analysis of the TFBS's in the promoter region of *CYP1A1* gave an inconsistent picture. For the DREs, the PWMs acquired from Jaspar and Transfac lead to a strong reduction in TF binding affinity but short motifs (AhR/Arnt from Jaspar and TCF4 from Transfac) biased the scores and consequently the reduction in binding affinity. The usage of the AhR/Arnt profile from Transfac for further analysis was also questionable because an analysis by Rahmann et al. (2003) certified the Transfac database disturbingly poor quality. For the DREs

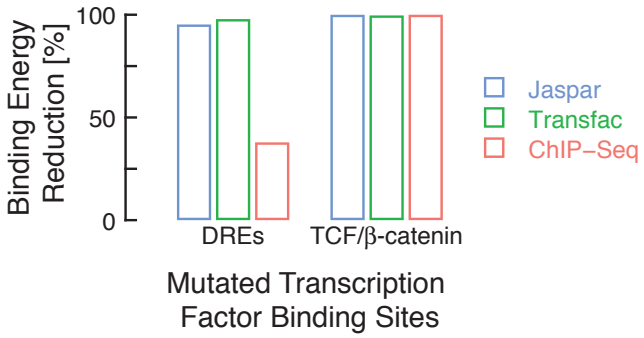


Figure 3.7: Reduction of binding affinity in response to point-mutations within the core binding sequences. The reduction of binding energy is given in percent for the DREs and the TCF/β-catenin TFBS's based on data from Jaspar, Transfac and ChIP-Seq.

the ChIP-Seq motif remained which however only reduced TF binding affinity for a mutated binding site by 40 %. A halved binding affinity does not necessarily mean that no TFs would bind to that site anymore. In contrast, one would expect that approximately one of two TFs is still able to bind to this binding site on average. Hence, one would expect to still observe an inducibility of a promoter whose binding sites are all mutated. Stimulating a promoter construct with no wild type binding sites with TCDD did not confirm this expectation (Figure A.1).

This lead me to conclude that binding to these mutated binding sites is negligible and that the estimates from the PWMs were inaccurate or at least flawed. Therefore, I treated the associations of TFs to the DREs and the TCF/β-catenin binding sites as all-or-none binding events, i.e., there is TF binding to the wild type binding sequence and there is no binding to any mutated binding site sequence.

3.3 Cooperativity between C- and D-DREs in the human *CYP1A1* promoter

AhR and Wnt/β-catenin signaling pathways converge on four DREs in the promoter region of *CYP1A1*. In recent work it was shown that multiple TFBS's can give rise to cooperative interactions between them and thus establish complex promoter logics (Casanovas et al., 2014; Giorgetti et al., 2010; Hermsen et al., 2010; Korenčič et al., 2012). To investigate the cooperative behavior between the binding sites in the *CYP1A1* promoter, we created promoter mutants in which the C- and the D-DRE

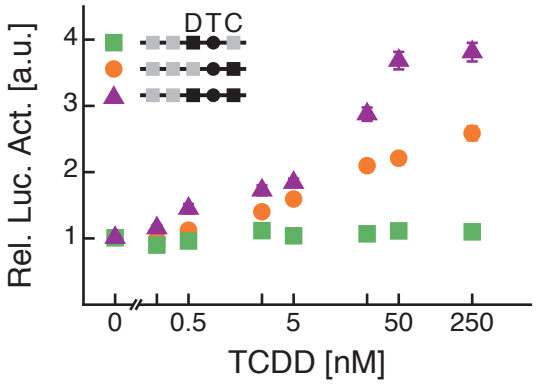


Figure 3.8: Cooperativity between TFs binding to C- and D-DREs on the natural promoter. Relative luciferase activity of three natural promoter constructs over a series of TCDD concentrations. On the three promoter constructs only the following TFBS combinations are present as WT sequence while the others are inactivated through point-mutation: square: TCF/β-catenin and D-DRE; circle: TCF/β-catenin and C-DRE; triangle: TCF/β-catenin, C- and D-DRE.

binding sequences were systematically point-mutated to disable their function (cf. Figure 2.1). Transfection in 55.1c cells and subsequent stimulation with increasing concentrations of TCDD resulted in the luciferase activities seen in Figure 3.8. The first reporter construct had only one functional, non-mutated DRE (D-DRE) and displayed hardly any induction by TCDD. In contrast, a construct with only the C-DRE binding site available for TF binding resulted a robust 2.5-fold increase in luciferase activity at higher amounts of TCDD. Finally, for a construct where both the C-DRE and the D-DRE were present, we observed an almost 4-fold increase in luciferase activity.

Thus, while the D-DRE alone did not change gene induction by TCDD, it cooperated with the C-DRE in inducing expression demonstrating a more-than-additive cooperativity between the interacting TFs that bind to these DREs.

3.4 Cooperativity and dual signal integration in synthetic promoters

The cooperative behavior between the TFs that bind to the C- and D-DRE motivated me to investigate their interaction in more detail. In particular I asked how the cooperative interaction is affected by different binding site sequences or distances

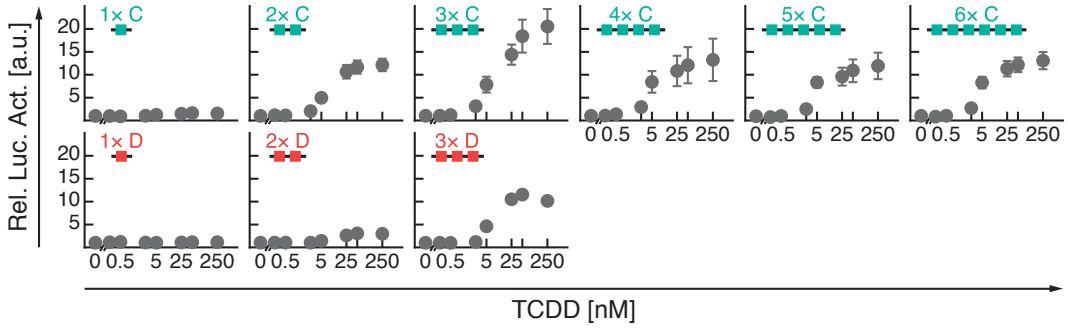


Figure 3.10: Concentration series of the synthetic promoter constructs. Relative luciferase activity of nine synthetic promoter construct (1x-6x C-DRE and 1x-3x D-DRE) for increasing TCDD (gray circles). Error bars represent one standard deviation of 6–10 biological replicates.

between the binding sites. Even though both C- and D-DRE share an equal core binding sequence (`5'-GCGTG-3'`), their flanking regions are highly different (cf. Figure 2.1) which lead to highly different binding affinities (Figure 3.9). For all three data sources (Jaspar, Transfac and ChIP-Seq) the TF binding to the C-DRE had a higher affinity than the one binding to the D-DRE.

For a systematic analysis of these DREs and their interactions with respect to changing promoter response, we constructed synthetic promoter mutants holding only the C- and D-DREs. The library of synthetic promoter mutants consisted of six constructs with one to six copies of the C-DRE TFBS sequence and three constructs with one to three copies of the D-DRE TFBS sequence (Figures **??** and 3.10). It should be noted that the distance of 19 bp between two TFBS's in the C-DRE and the D-DRE constructs was smaller than the mean TFBS distance of 260 bp in the wild type *CYP1A1* promoter. Additionally, the sequences between the TFBS's were chosen not to be AhR-responsive. These nine constructs were then separately transfected into 55.1c cells and subsequently treated with increasing concentrations of TCDD for 24 h. The resulting luciferase activity data was preprocessed as described in Chapter 2.2 and is shown in Figure 3.10. The reporter constructs containing either one C-DRE or one D-DRE exhibited no significant induction of reporter

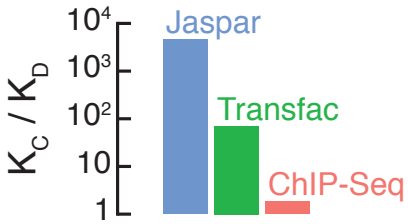


Figure 3.9: Binding affinity difference between C- and D-DRE. The quotient between the association constants of TFs to the C-DRE K_C and the D-DRE K_D is shown for data from Jaspar, Transfac and ChIP-Seq.

gene expression for increasing TCDD concentrations. In the case of the C-DRE constructs, I observed a maximal induction of 20-fold for the reporter construct with three present DREs. Between the C- and the D-DRE constructs, a higher induction could be observed in the C-DRE constructs than in the corresponding D-DRE construct. For example, the maximal fold-change for the double C-DRE construct was 12 whereas it was only 3 for the double D-DRE construct. It was reassuring that this difference is in line with the difference of binding affinity as estimated from the high-throughput data (cf. Figure 3.9).

In summary, I observed a dose-dependent induction for all synthetic promoter constructs while the C-DREs could be stronger induced by TCDD than the D-DREs. Since the induction of the C-DREs decreased with four or more TFBS's, it appeared that adding more TFBS's to the construct did not add to the inducibility of the promoter but conversely decremented its induction.

3.4.1 Modeling the synthetic promoters

The counterintuitive results of the synthetic promoter constructs called for the help of a mathematical model to unravel the underlying mechanisms. Thus, I constructed a two-part model. The first part consisted of a small signaling model outlining the crosstalk between the AhR and the Wnt/β-catenin signaling pathways as well as the formation of the TFs complex acting on the DREs within the promoter region of the *CYP1A1* gene. The second part comprised a thermodynamic model relating the fold-change expression of the promoter to the interactions between the TFs associating with the various binding sites within the *cis*-regulatory region.

Model of the AhR and Wnt/β-catenin signaling pathways

Exposure to TCDD leads to its intrusion into the cells of hepatic lobules. There, the presence of such exogenous chemicals triggers the activation of the AhR signaling pathway through binding to AhR. The ligand-receptor complex then translocates into the nucleus where it recruits Arnt. As outlined in Figure 3.11, I assumed for simplicity that this step is carried out by just one reaction. As shown in Chapter 3.1, β-catenin binds to AhR and thus is a member of the TCDD/AhR/Arnt TF complex that associates with the DREs in the *CYP1A1* promoter. Therefore, I assumed that only the complex containing β-catenin associates with the DREs. This process is

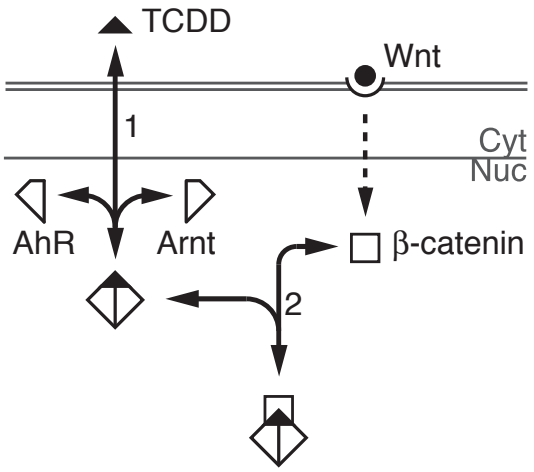


Figure 3.11: Minimal model of the crosstalk between the AhR and the Wnt/β-catenin signaling pathways. Reaction 1 summarizes the formation of the TCDD/AhR/Arnt complex while reaction 2 describes binding of β-catenin to said complex.

qualitatively captured by

$$T + R + N \overset{1}{\rightleftharpoons} A \tag{3.1}$$

$$A + B \overset{2}{\rightleftharpoons} F\,. \tag{3.2}$$

Herein T represents TCDD while R, N, A, B and F depict AhR, Arnt, the whole complex, β-catenin and the fully formed TF, respectively. By assuming that the system operates in thermodynamic equilibrium (which was also the major assumption for the upcoming thermodynamic model of the *CYP1A1* promoter), the equations

$$k_1^+\, T\, R\, N = k_1^-\, A \tag{3.3}$$

$$k_2^+\, A\, B = k_2^-\, F \tag{3.4}$$

describe the reaction system of Equation 3.2, where k_i^+ and k_i^- are the forward and backward reaction rates. Additionally, the total concentration of TCDD T_{Tot} and β-catenin B_{Tot} in the system shall be conserved such that

$$A + T + F = T_{\mathrm{Tot}} \tag{3.5}$$

$$B + F = B_{\mathrm{Tot}}\,. \tag{3.6}$$

With the definition of the dissociation constant $K_i = k_i^- / k_i^+$, the concentration of the TCDD/AhR/Arnt complex is defined by Equations 3.3 and 3.5 as

$$A = \underbrace{\frac{R\,N}{K_1 + R\,N}}_{\alpha} (T_{\mathrm{Tot}} - F)\ . \tag{3.7}$$

Since neither the concentrations of AhR (R) and Arnt (N) nor the dissociation constant K_1 were measured, I lumped them into the variable α. With the help of Equations 3.4, 3.6 and 3.7 the concentration of the TF F can now be calculated from

$$F^2 - F\left(T_{\mathrm{Tot}} + B_{\mathrm{Tot}} + \frac{K_2}{\alpha}\right) + T_{\mathrm{Tot}}\,B_{\mathrm{Tot}} = 0 \tag{3.8}$$

to

$$F = \frac{1}{2}\left(B_{\mathrm{act}}\,B_{\mathrm{Tot}} + T_{\mathrm{Tot}} + \frac{K_2}{\alpha} - \sqrt{\left(-B_{\mathrm{act}}\,B_{\mathrm{Tot}} - \frac{K_2}{\alpha} - T_{\mathrm{Tot}}\right)^2 - 4\,B_{\mathrm{act}}\,B_{\mathrm{Tot}}\,T_{\mathrm{Tot}}}\right)\ . \tag{3.9}$$

Furthermore, since the total concentration of β-catenin was not measured, I added an additional scaling factor B_{act} representing the percentage of active β-catenin. While this factor represents the modulation of a β-catenin inhibitor, B_{Tot} was estimated from the experimental data. From the conservation relations it follows that F is always smaller than T_{Tot}. Thus, the solution of the quadratic equation with the positive root is not feasible.

The analytic expression for the TF concentration F now let me utilize it as an inputs into the various thermodynamic models I will derive in the following chapters.

Thermodynamic models of the synthetic promoter constructs

Thermodynamic models describe the binding of TFs to binding sites in the promoter region of target genes. The library of synthetic promoter constructs contained reporters with one to six equivalent C-DREs and one to three equivalent D-DREs (cf. Figure **??**). Instead of setting up nine different formulations of the matrices within the matrix-based thermodynamic modeling frame work (cf. Chapter 2.4), I used

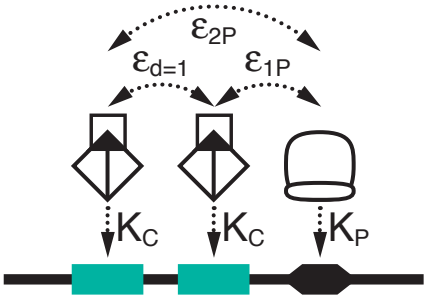


Figure 3.12: Depiction of the parameters of the thermodynamic model describing the double C-DRE promoter construct. K_i represents the association constants between i and the DNA, ϵ_{iP} describes the adhesive binding energy between the TF binding to the i-th C-DRE (as counted from the RNAP binding site) and the RNAP, and $\epsilon_{d=i}$ represents the cooperative binding between two TFs over the distance i.

the architectural symmetries of the constructs. Since for example all C-DRE constructs only harbor different amounts of equivalent binding sites, it was sufficient to generate one set of the $\boldsymbol{L}$, $\boldsymbol{s}$, $\boldsymbol{t}$ and $\boldsymbol{b}$ matrices for the largest construct. By selectively setting the association constants to zero the model formulations of all other constructs with fewer binding sites could be produced. But since I showed previously that the association constant of the TF binding to the C-DRE was significantly larger than the one of the TF binding the D-DRE (cf. Figure 3.9), different formulations for the C-DRE and the D-DRE constructs were necessary. The presence of equivalent binding sites within the two classes of constructs allowed me to further reduce the parameter space by assuming that the respective association constants were equal too, i.e. all TFs associate to binding site within the C-DREs constructs with K_C whereas TFs bind to the TFBS's within the D-DRE constructs with K_D. The constant distance of 19 bp between the TFBS's in the synthetic constructs furthermore permitted that TFs binding to equidistant TFBS's exhibit the equivalent binding energy. TFs that are direct neighbors for example shared one equal cooperative binding energy $\epsilon_{d=1}$ while TFs communicating over one intermediate TF shared the cooperative binding energy $\epsilon_{d=2}$ and so on. Based on the matrix-based thermodynamic modeling framework as introduced in Chapter 2.4 and the aforementioned assumptions, I exemplify the derivation of the models for the synthetic promoter constructs on the double C-DRE construct (Figure 3.12). Since only states with a bound RNAP lead to transcription the transcription vector was

$$\boldsymbol{t}^T = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} . \tag{3.10}$$

The position matrix followed to

$$\boldsymbol{L}^T = \begin{bmatrix} 1 & 1 & 1 & 1 & [F] & [F] & [F] & [F] \\ 1 & 1 & [F] & [F] & 1 & 1 & [F] & [F] \\ 1 & [P] & 1 & [P] & 1 & [P] & 1 & [P] \end{bmatrix} \tag{3.11}$$

while the equilibrium constants vector was

$$\boldsymbol{b} = \begin{bmatrix} 1 \\ K_P \\ K_C \\ K_P\, K_C\, \zeta_1([P],[F]) \\ K_C \\ K_P\, K_C\, \zeta_2([P],[F]) \\ K_C^2\, \zeta_1([F],[F]) \\ K_P\, K_C^2\, \zeta_1([P],[F])\, \zeta_2([P],[F])\, \zeta_1([F],[F]) \end{bmatrix} \equiv \begin{bmatrix} \vdots \\ \\ b_{PC_1} \\ \vdots \\ b_{PC_2} \\ b_{C_{d=1}} \\ b_{PC_1C_2} \end{bmatrix} . \tag{3.12}$$

Herein $\zeta_i([F],[F]) = \exp(-\epsilon_{d=i}/k_BT)$ and $\zeta_i([P],[F]) = \exp(-\epsilon_{iP}/k_BT)$ depending on the distance i between the two C-DREs and between the RNAP and the i-th C-DRE, respectively. The concentration of the TF was given by $[F]$ as calculated from Equation 3.9. The complete formulation of the matrices describing the nine synthetic C- and D-DRE promoter constructs can be found in Appendix A.1. The binding probability was calculated with Equation 2.25 to

$$p_{\text{bound}}([P],[F],[F]) = \frac{\overbrace{K_P\,[P]}^{q_P} + \overbrace{b_{PC_1}\,[P]\,[F]}^{q_{PC_1}} + \overbrace{b_{PC_2}\,[P]\,[F]}^{q_{PC_2}} + \overbrace{b_{PC_1C_2}\,[P]\,[F]^2}^{q_{PC_1C_2}}}{1 + K_P\,[P] + \underbrace{K_C^2\,[F]^2}_{q_{C_1}+q_{C_2}} + (b_{PC_1} + b_{PC_2})\,[P]\,[F] + \underbrace{b_{C_1C_2}\,[F]^2}_{q_{C_1C_2}} + b_{PC_1C_2}\,[P]\,[F]^2} . \tag{3.13}$$

Since, the measurements of the luciferase activity of the synthetic reporter constructs were not absolute but relative to the basal reporter activity of one construct, each construct was normalized to its specific basal activity (cf. Chapter 2.2.1). To reflect these normalizations with the model the binding probability too had to be

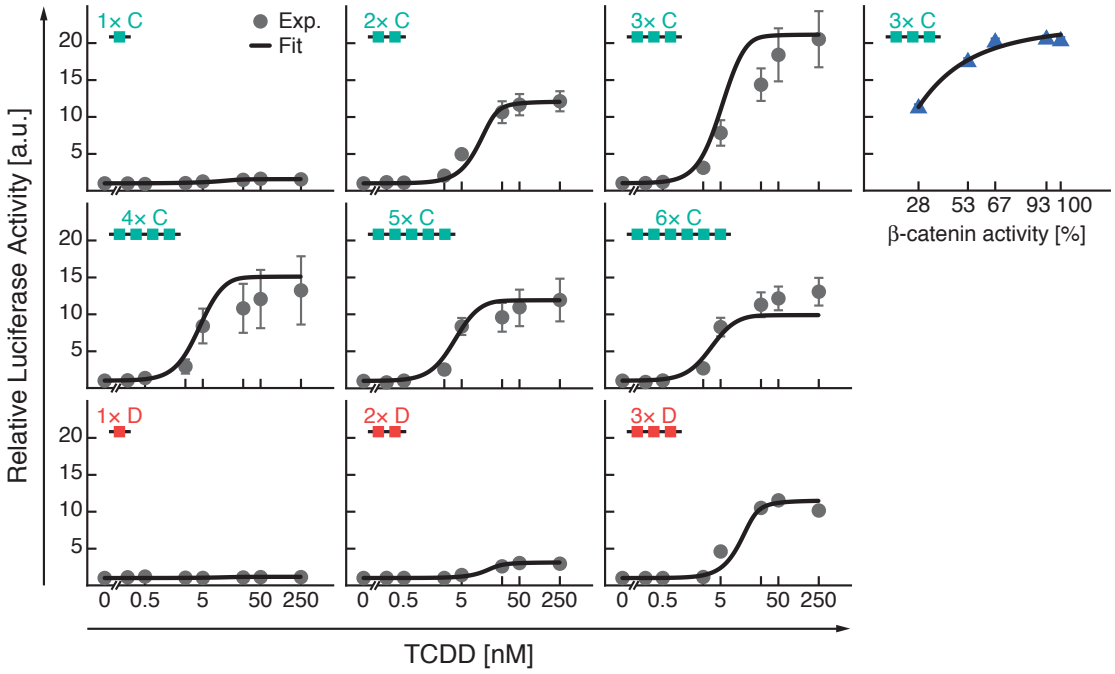


Figure 3.13: Mathematical model could explain the synthetic promoter constructs. Relative luciferase activity of nine synthetic promoter construct (1x-6x C-DRE and 1x-3x D-DRE) for increasing TCDD (gray circles) as well as one β-catenin titration series (blue triangles) was measured. The β-catenin titration was measured at 250 nM TCDD. Error bars represent one standard deviation of 6–10 biological replicates. Black curves depict the estimations of the mathematical model.

normalized to the basal transcription rate such that

$$\phi = \frac{p_{\text{bound}}([P],[F],[F])}{p_{\text{bound}}([P],[F]=0,[F]=0)} \,. \tag{3.14}$$

Since the experimental setup allowed to measure neither the concentration of the RNAP nor its association constant directly q_P was estimated from the data.

3.4.2 Model explains cooperativity in synthetic promoters

As I now established the mathematical model, I used the strategy outlined in Chapter 2.6 to estimate the 17 free parameters of the models describing the C- and D-DREs from the data. In Figure 3.13 it can be seen that the model was able to describe the data in most quantitative detail (compare black solid lines with gray and blue data points). In order to be able to identify the impact of Wnt/β-catenin signaling, i.e.,

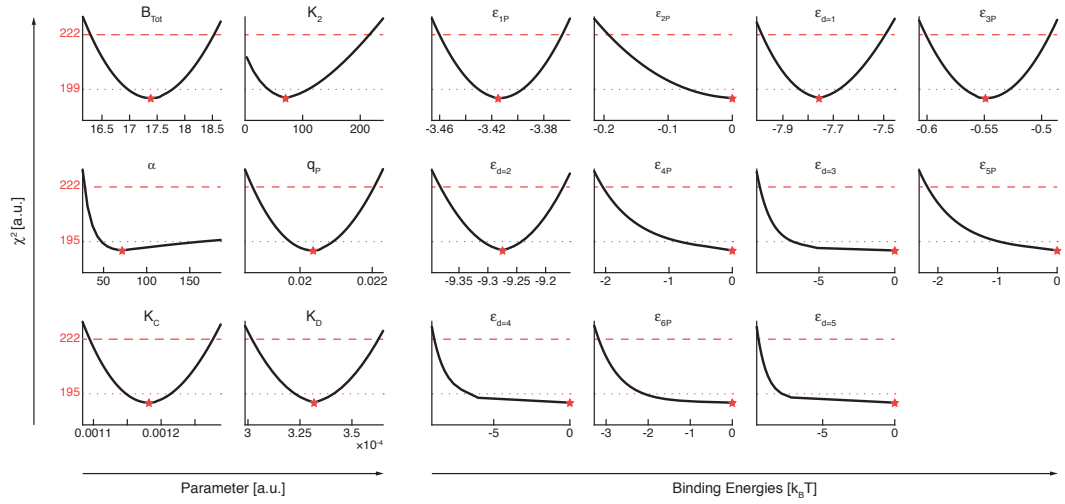


Figure 3.14: All parameters of the thermodynamic model were identifiable within the biologically feasible limits. The black lines display the profile likelihood versus the parameter values. The optimal parameter values as determined by the model fitting procedure are display as red stars. The point-wise as well as the simultaneous threshold for the confidence intervals are represented by red dotted and dashed lines, respectively.

the parameter B_{tot}, I added a β-catenin titration series to the data set from which the model parameters were estimated. The mathematical model was able to reproduce the strong sigmoidal response. All constructs displayed an almost switch-like behavior if TCDD concentration was increased. Even though the data of the C- and the D-DRE constructs differed significantly, they could be explained by only one differing parameter that models the binding affinity of the TFs to the DRE sequences. In particular, as predicted by the bioinformatic analysis (cf. Figure 3.9), binding affinity of the TFs to the C-DRE is four times stronger than that to the D-DRE (cf. Table A.1). The confidences of the parameters were established with the help of the profile likelihood method outlined in Chapter 2.5. Applying a point-wise confidence threshold, i.e., confidence intervals hold individually for each parameter, I found that all parameters had a clear minimum and thus were identifiable within the biologically feasible limits (Figure 3.14). Ten parameters had relatively small confidence intervals and were significantly larger than zero while the remaining seven parameters (ϵ_{2P}, ϵ_{4P}, ϵ_{5P}, ϵ_{6P}, $\epsilon_{d=3}$, $\epsilon_{d=4}$ and $\epsilon_{d=5}$) were indistinguishable from zero, i.e., the model was able to fit the data without these interactions.

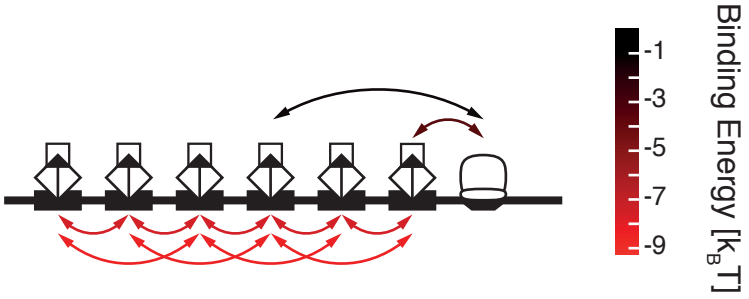


Figure 3.15: Cooperative binding dominated the synthetic promoter constructs. Binding energies resulting from fits displayed on the promoter construct holding six copies of the C-DRE. Arrows depict significant binding events. Their colors represent binding strength where lower values represent stronger association. All binding events were present in the model. Those equal or close to zero are not depicted as arrows.

Cooperativity limited to nearby TFs

The estimated set of parameters (cf. Table A.1) now aided the understanding of the observed induction behaviors of the synthetic reporter constructs (Figure 3.15). I found that the adhesive binding energy to the RNAP was only significant for the TF binding to the DRE closest to the RNAP binding site (ϵ_{1P} was approximately five times larger than ϵ_{3P}). Furthermore, only those parameters describing short range cooperative binding ($\epsilon_{d=1}$ and $\epsilon_{d=2}$) were estimated to be larger than zero. Thus, cooperative binding of TFs occured only between direct neighbors or over one intermediate TF. Longer distance adhesive binding between TFs and RNAP was either zero or close to zero, as was longer distance cooperative binding between TFs.

To confirm these modeling results, we created four additional synthetic promoter constructs in which two C-DREs were placed at distances 19 bp, 49 bp, 156 bp and 292 bp from each other (Figure 3.16). These constructs were again transfected into 55.1c cells and exposed to increasing concentrations of TCDD. Here, the maximal fold-change of 12.5-fold could be observed in the construct with the smallest distance of 19 bp between the C-DREs. Note that these separate experiments were in good agreement with the induction behavior as observed in Figure 3.13. With increasing distance the maximal induction decreased to 11-fold for a distance of 49 bp and to 2.5-fold for the distances 156 bp and 292 bp. Thus, inducibility decreased with increasing distance between the TFBS's suggesting that, as predicted

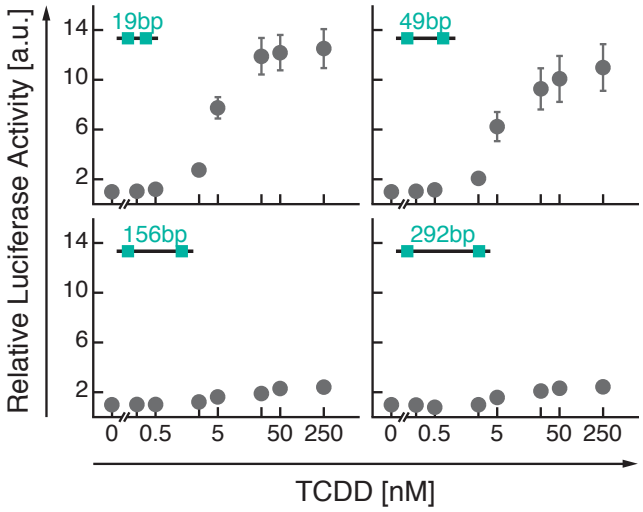
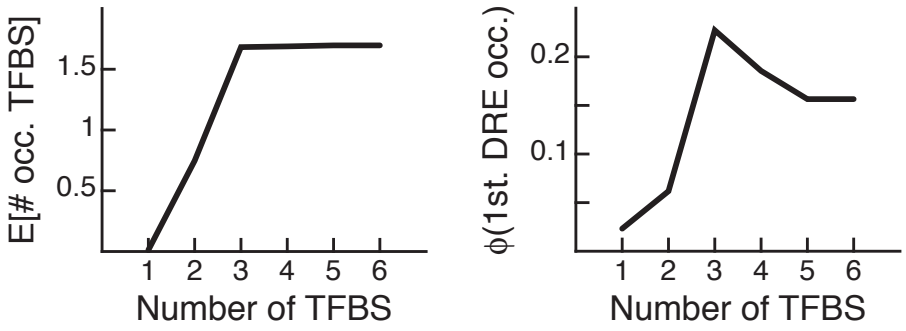


Figure 3.16: Promoter induction decreased with distance. Measurements of the relative luciferase activity over increasing TCDD concentrations of four synthetic constructs with 19 bp, 49 bp, 156 bp and 292 bp between two C-DRE TFBS's is shown.

by the model parameters, long distance cooperativity was limited.

Sequestration reduced transcriptional induction

A counterintuitive aspect of the synthetic promoter data was that adding more DREs to the promoter did not necessarily lead to increased transcriptional output. Adding more than three DREs rather decreased activity (Figure 3.13). Interestingly, the mathematical model could accurately describe this behavior. In the following I will hypothesize on the underlying mechanism. Since only the most proximal DRE significantly interacted with the RNAP, transcription was only initiated if the TCDD/AhR/Arnt/β-catenin TF complex was bound to the first DRE. Furthermore, the DREs only interacted over short distances, i.e., with the neighboring DRE and the DRE next to its neighbor. So, for the case when two DREs were present, the second DRE could interact with the first DRE and thus increase transcriptional output. Similarly, if three DREs were present both the second and the third DRE could interact with the first DRE and further increase the transcriptional output. However, if the construct harbored four DREs, the forth DRE could not directly interact with the first DRE because the model parameters were zero for long range interactions. Since only pairwise interactions could occur, there were states where the second, third and forth DRE interacted with each other but none of them interacted



(a) Expected number of occupied binding sites

(b) Probability that first DRE is occupied

Figure 3.17: Sequestration reduced transcriptional induction. **(a)** Expected mean of occupied binding sites over number of sites present in construct. **(b)** Probability that the first DRE is bound by a TF plotted over the number of present TFBS's.

with the first DRE. This reduced transcriptional output and could thus be viewed such that the forth, fifth and sixth DREs sequestered interaction partners for the first DRE.

This hypothesis was also supported by the calculation of the average number of occupied binding sites as well as the probability that the first TFBS was bound by a TF. The expected mean of occupied binding sites was calculated by weighting the terms in the binding probability with the number of TFs in said terms. For the example of the synthetic construct with double C-DREs introduced in Chapter 3.4.1 the mean expectancy of the number of occupied binding sites follows from Equation 3.13 to

$$\mathbb{E}(\text{2x C-DRE}) = \frac{0 \cdot q_P + 1 \cdot q_{C_1} + 1 \cdot q_{C_2} + 1 \cdot q_{PC_1} + 1 \cdot q_{PC_2} + 2 \cdot q_{C_1C_2} + 2 \cdot q_{PC_1C_2}}{1 + q_P + q_{C_1} + q_{C_2} + q_{PC_1} + q_{PC_2} + q_{C_1C_2} + q_{PC_1C_2}} . \quad (3.15)$$

Figure 3.17a shows the progression of separately calculated expectancy values for each number of TFBS's present in the synthetic constructs. While no TFs were bound on average if only one C-DRE was present (explaining the low induction of the 1x C-DRE construct in Figure 3.10), one TFBS was occupied in the 2x C-DRE construct. The maximum number of occupied TFBS's was reached in the triple C-DRE construct. Adding more DREs thus did not increase the mean number of

bound binding sites.

The probability that the first DRE and the RNAP binding site were occupied was calculated by summing up all states in which the first DRE and the RNAP are bound and dividing this by all possible states. For the double C-DRE of Chapter 3.4.1 this is

$$p_{\text{bound}}([P],[F],[F]=0) = \frac{q_{PC_1} + q_{PC_1C_2}}{1 + q_P + q_{C_1} + q_{C_2} + q_{PC_1} + q_{PC_2} + q_{C_1C_2} + q_{PC_1C_2}} . \quad (3.16)$$

Figure 3.17b shows the results obtained from separate calculations of these probabilities for each of the six constructs. The probability that the first and thus the strongest cooperator with the RNAP was bound to its binding site presented a peak for the triple C-DRE construct. Thus, the same number of TFs may be distributed over four or more TFBS's resulting in states where the first TFBS was unoccupied and induction was reduced.

In summary, while the number of occupied binding sites remained constant for three or more DREs, the present TFs had more possibilities to bind to the promoter leading to a decreased probability that the first DRE was occupied. And since the TF that was bound to the first DRE interacted most strongly with the RNAP this lead to decreased induction of the 4x to 6x C-DRE constructs.

3.4.3 Prediction of dual signal integration by the synthetic promoters

Up until now I established a modeling framework that could describe the response to TCDD stimulation of the synthetic constructs in most quantitative detail. Apart from the β-catenin titration series of the triple C-DRE construct (cf. Figure 3.13) the impact of the Wnt/β-catenin signaling and the influence of its crosstalk with the AhR pathway on promoter induction has not yet been studied.

The mathematical model was trained on data from 55.1c cells that have a constitutively active form of β-catenin. With the help of the parameter B_{act} (cf. Equation 3.9) that represented the β-catenin activity the inhibition of β-catenin by iCRT3 could be predicted. In the left panels of Figure 3.18, I plotted the predicted response of the model to all possible combinations of β-catenin and TCDD levels for constructs

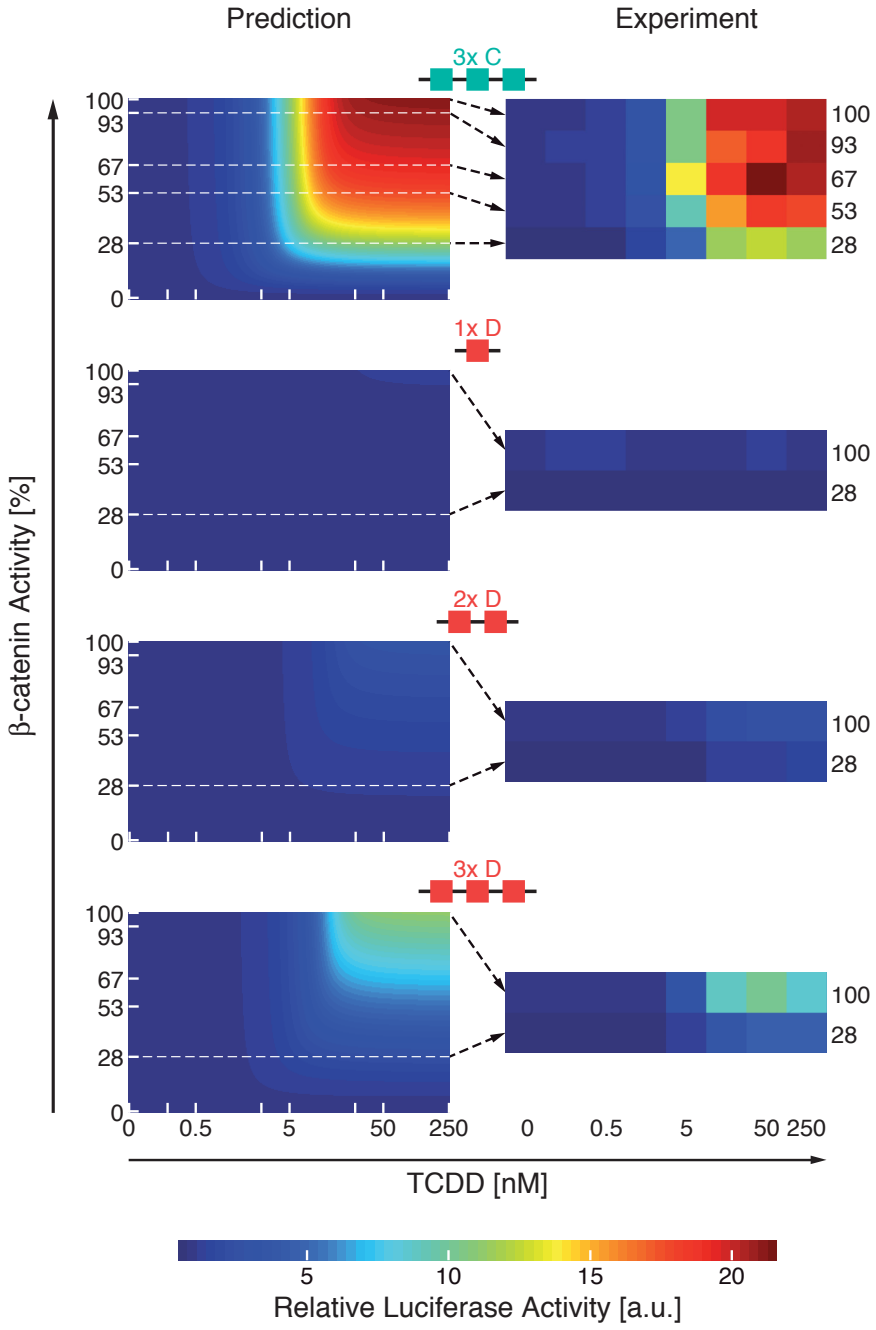


Figure 3.18: Mathematical model correctly predicted and experimental data confirmed an AND-gate relationship between AhR and Wnt/β-catenin signaling pathways for one triple C-DRE construct and three D-DRE constructs with one to three TFBS's. Left: Predictions of the promoter activity by the mathematical model through variation of the β-catenin activity parameter B_{act}. Right: Corresponding measurements of the relative luciferase activity where iCRT3 modulates β-catenin activity over increasing TCDD concentrations.

holding triple C-DREs and one to three D-DREs. The model predicted that the *CYP1A1* promoter requires both AhR and Wnt/β-catenin signaling pathways to be active in order to induce expression. With both 100 % β-catenin and 250 nM TCDD the relative luciferase activity was 21. By calculating the row and column means of the relative luciferase activity I could infer if the promoter was inducible for a given β-catenin activity and TCDD concentration, respectively. Setting the threshold of mean luciferase activity to 3 resulted in a β-catenin limit of 12 % and a TCDD concentration limit of 2.5 nM. Thus, for low β-catenin activity high doses of TCDD were not sufficient to activate transcription in the triple C-DRE construct. Similarly, full β-catenin activity did not result in a promoter response if only low TCDD concentration were present. Additionally, if the threshold (12 % β-catenin activity and 2.5 nM TCDD) was crossed, small changes in stimulation lead to large changes in the promoter response. This ultra-sensitivity resembled the behavior of a logic AND-gate. For the D-DRE constructs a similar, but much less pronounced response could be observed. The construct harboring only one D-DRE was not inducible at all with relative luciferase activities ranging from 1 to 1.15. Similarly, the double D-DRE construct only displayed a relative luciferase activity above 3 for 98.5 % β-catenin activity and 200 nM TCDD. The triple D-DRE construct on the other hand reached a maximal relative luciferase activity of 11 but still high β-catenin activity (58 %) and TCDD concentrations (17 nM) were necessary to induce a promoter response. An alleviated AND-gate could be observed for this construct.

To test the predictions made by the mathematical model experimentally, we systematically varied the β-catenin activity with the inhibitor iCRT3 and increased the TCDD concentration. Afterwards we measured the relative luciferase activity for the aforementioned synthetic constructs (Figure 3.18 right). For the triple C-DRE construct it can be observed that a decrease in β-catenin activity resulted in a lower luciferase activity. Similarly, as already observed in the synthetic constructs that were only stimulated with TCDD (cf. Figure 3.10), a lower TCDD concentration lead to a decrease in *CYP1A1* expression. Similar observations could also be made for the constructs containing D-DREs. Even though a full inhibition of β-catenin activity and thus a complete shutdown of *CYP1A1* induction was experimentally not accessible, this data was in quantitative agreement with the model predictions.

Thus, the ultra-sensitive AND-gate behavior as predicted by the mathematical model was confirmed in the experimentally accessible range.

3.5 Cooperativity and dual signal integration in the human *CYP1A1* promoter

In the previous chapter I established that the mathematical model was indeed able to describe the cooperativity in well-controllable synthetic reporter constructs. The therein utilized approach was applied to analyze the human *CYP1A1* promoter in the following. Just as with the synthetic constructs, we created a promoter construct library containing constructs with 17 combinatorial combinations of point-mutations that inactivated specific TFBS's (Figures 2.1, 3.19 and A.2). Each of these mutant natural reporter constructs was subsequently transfected into mouse hepatoma cells of line 55.1c and treated with increasing concentrations of TCDD for 24 h. The resulting relative luciferase activity data was preprocessed as described in Chapter 2.2 and exhibited a less obvious behavior than in the synthetic constructs (gray circles in Figure 3.19). For example, the reporter construct containing all binding sites except the F-DRE (EDTC construct) displayed the strongest induction of 7-fold upon TCDD treatment, even stronger than the 6-fold induction of the wild type construct that harbored no inactivated TFBS's. Additionally, the constructs DTC, FTC, EDT, FET, FDTC, FEDC and FEDT display almost the same 4-fold inducibility although they contain different types and numbers of binding sites. Finally, it was unexpected that the constructs TC and ET presented with a 2.5-fold inducibility while the DT and FT constructs are hardly inducible at all.

To disentangle the effects of the different binding sites and analyze if the behavior of the promoter can be understood by the interactions between the binding sites, I anew turned to mathematical modeling.

3.5.1 Model of the natural promoters

Building on the approach introduced in Chapters 2.4 and 3.4.1, I populated the matrices $\boldsymbol{L}$, $\boldsymbol{s}$, $\boldsymbol{t}$ and $\boldsymbol{b}$ to reflect the structure and mutation patterns of the natural reporter constructs (see Appendix A.1) to calculate the normalized binding probabilities for each of the present 18 natural reporter constructs.

In order to reduce the dimensionality of the parameter space the following assumptions were taken into account. The association constants of the C- and D-DRE

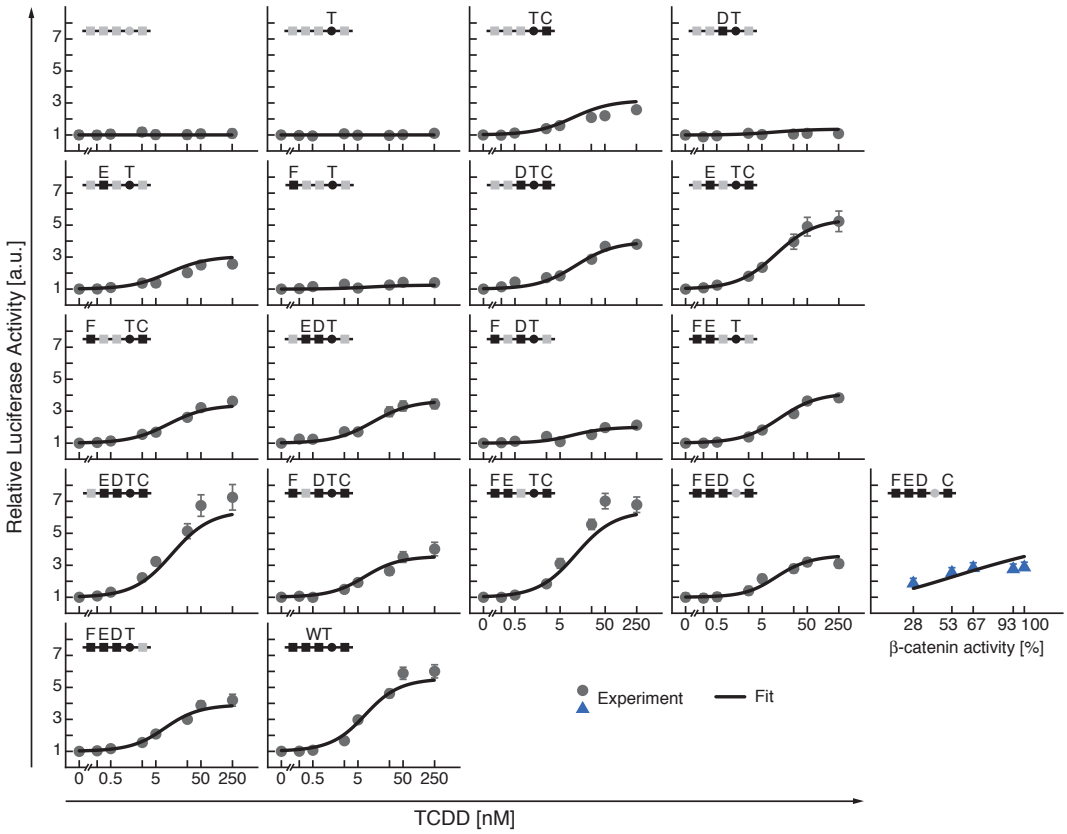


Figure 3.19: Mathematical model could explain natural promoter induction. Relative luciferase activity of 18 natural reporter constructs for increasing TCDD concentrations (gray circles) as well as one β-catenin titration series (blue triangles). The β-catenin titration series was measured at 250 nM TCDD. Error bars represent one standard deviation of 4–17 biological replicates. The black curves show the estimations of the mathematical model.

were already estimated in the model of the synthetic reporter constructs (cf. Table A.1). Therefore they could be transferred to the model of the natural promoter. Within the synthetic constructs I furthermore saw that cooperative binding only occured between direct or next neighbors. Consequently, the long-distance interactions (i.e. between C-DRE and E-DRE, C-DRE and F-DRE, and TCF/β-catenin and F-DRE) were eliminated from the model. As elaborated on in Chapter 3.2, binding to TFBS's occured as all-or-none binding processes, i.e., mutation of a TFBS hindered binding of a TF. Hence, if a binding site was inactivated through a mutation in the construct I removed the mutant TFBS from the model by setting the corresponding association constant between TF and TFBS to zero. This matrix-based thermodynamic representation of the natural *CYP1A1* promoter occupancy

was subsequently joined with the signaling model as derived in Chapter 3.4.1.

Estimation of the remaining 19 parameters from the entire data including a β-catenin titration series resulted in an agreement of the data and the model (black lines in Figure 3.19). While the strong inductions of the constructs EDTC and FETC were slightly underestimated, the induction of the TC constructs was slightly overestimated.

An interesting aspect of the experimental data and the model was that the induction of the natural promoter (Figure 3.19) showed a less pronounced sigmoidality, i.e., not such a strong switch-like behavior as compared to the synthetic promoters (Figure 3.13).

3.5.2 Cooperativity in the natural promoters

To understand in more detail how the difference in the sigmoidality of the stimulus-response curves could arise, I investigated the parameter values (cf. Table A.2). By applying the aforementioned profile likelihood method (cf. Chapter 2.5) I analyzed which parameters could robustly be estimated from the experimental data. I found that 18 of the 19 parameters showed a clear minimum and were thus identifiable in the biological feasible limits by the data when point-wise confidence thresholds were applied (Figure 3.20). Two of the parameters (q_P and ϵ_{DP}) were indistinguishable from zero, i.e., irrelevant for the model to describe the data. Only the cooperative binding energy between the D- and the E-DRE (ϵ_{ED}) was practically non-identifiably and thus could not be reliably determined from the data.

When I compared the adhesive binding energies between the TFs and the RNAP with those from the synthetic promoters (cf. Tables A.1 and A.2), I found that they were generally higher in the natural promoter constructs. Furthermore, they were relatively independent of the distance to the RNAP binding site (Figure 3.21). While the adhesive binding energy between the D-DRE and the RNAP (ϵ_{DP}) was zero, its cooperative interaction with the TCF/β-catenin TF (ϵ_{DT}) was the strongest. On the other hand, the strong adhesive binding energies of the C-DRE and the E-DRE with the RNAP (ϵ_{CP} and ϵ_{EP}) resulted also in strong cooperative interactions to the remaining TFBS's (ϵ_{TC}, ϵ_{DC}, ϵ_{ET}, ϵ_{ED} and ϵ_{FE}). But most striking was the strong influence of the TCF/β-catenin TFBS. It not only exhibited the strongest connection with the RNAP (ϵ_{TP}) but also communicated most strongly with its direct or next neighbors (ϵ_{TC}, ϵ_{DT} and ϵ_{ET}).

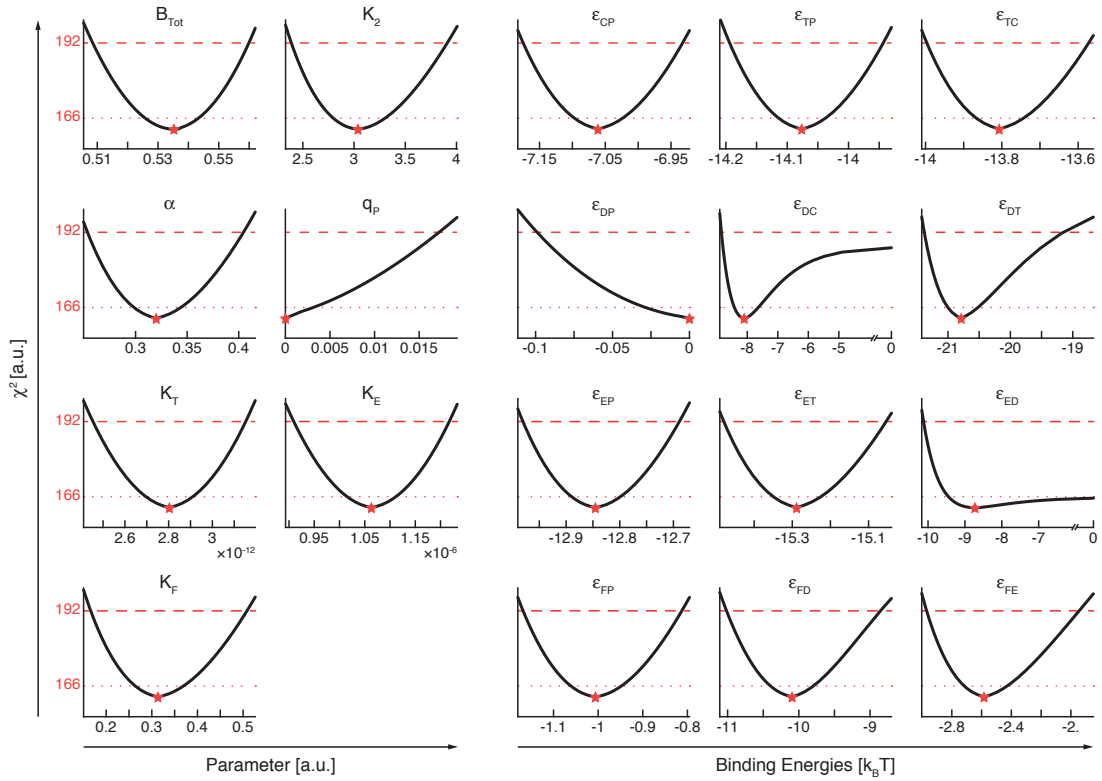


Figure 3.20: Most of the parameters of the mathematical model were identifiable. The black lines display the profile likelihood versus the parameter values. The optimal parameter values as determined by the model fitting procedure are display as red stars. The pointwise as well as the simultaneous threshold for the confidence intervals is represented by red dotted and dashed lines.

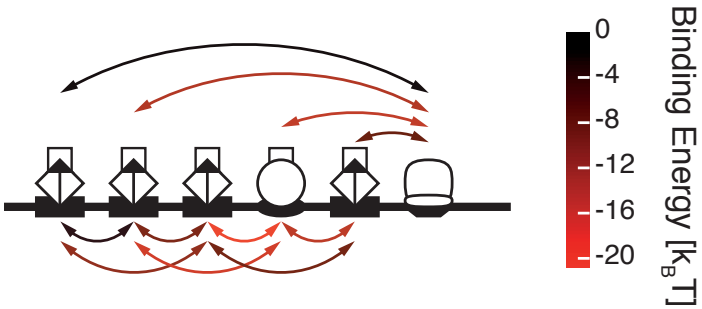


Figure 3.21: Binding between the RNAP and the TFs dominated the natural promoter. Binding energies resulting from fits displayed on the natural promoter construct. Arrows depict significant binding events. Their colors represent binding strength where lower values represent stronger association. Only short-range binding was present in the model. Those equal or close to zero are not depicted as arrows.

In summary, the experimental data of the natural reporter constructs depicted a diverse response to TCDD stimulations. With the help the mathematical model I could demonstrate the importance of TFBS targeted by the Wnt/β-catenin signaling pathway in the regulation of *CYP1A1* expression.

3.5.3 Prediction of dual signal integration by the natural promoters

Previously, I established that the mathematical model is able to quantitatively describe the response of multiple natural *CYP1A1* promoter constructs to TCDD stimulation. As with the synthetic constructs, I subsequently examined the predictive potential of the model regarding the integration of AhR and Wnt/β-catenin signaling into the promoter.

For this analysis I focused on three constructs: the TC construct harboring only wild type binding sequences for the C-DRE and the TCF/β-catenin TFBS, the FEDC construct in which only the TCF/β-catenin was mutated and most importantly, the wild type human *CYP1A1* promoter construct. The FEDC construct was chosen because it elucidates the crosstalk between AhR and Wnt/β-catenin signaling that occurs upstream of the promoter. Through variation of the model parameter B_{act} (cf. Equation 3.9) the influence of β-catenin activity was quantified. These predictions were compared to experimental data in which each construct was treated with different iCRT3 and TCDD concentrations (Figure 3.22). For the TC construct the model predicted a weak induction of 2.8 relative luciferase activity that was established with both 100 % β-catenin activity and 250 nM TCDD (Figure 3.22 top left). By using a threshold of 1.5 relative luciferase activity, I determined if the corresponding construct is inducible. In the same way as for the synthetic constructs, I calculated the row and column mean to identify an effect of the β-catenin activity and the TCDD concentration, respectively. Thus, the TC construct was inducible if β-catenin activity was above 73 % while TCDD concentration was at the same time larger than 25 nM. This was similar to what we observed in the experimentally measured reporter activity for this construct (Figure 3.22 top right). The model furthermore predicted a 3.5-fold induction for the FEDC construct if both TCDD and β-catenin were at high doses (Figure 3.22 middle left). The inducibility limits for this construct were 44 % β-catenin activity and 5 nM TCDD. Similarly, the corresponding experimental data set showed a maximal induction of 3.25 ± 0.5 relative luciferase activity confirming that regulatory crosstalk is possible independent of

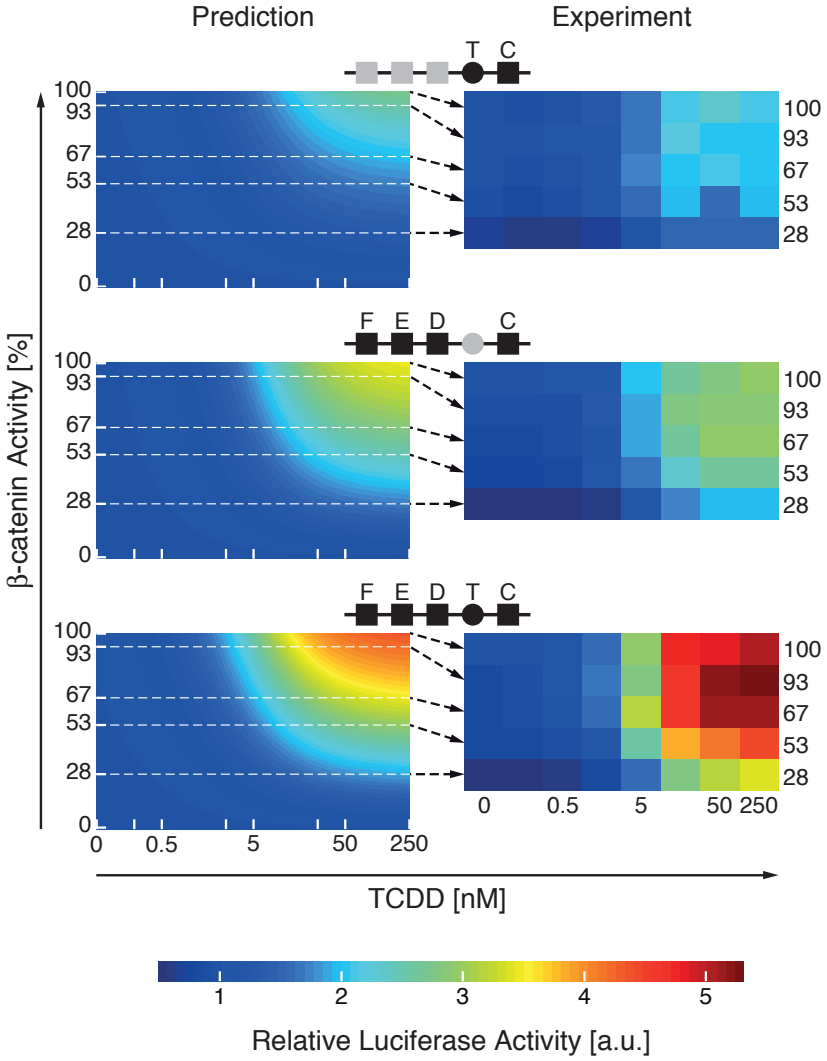


Figure 3.22: Mathematical model correctly predicted and experimental data confirmed an AND-gate relationship between AhR and Wnt/β-catenin signaling pathways for the TC and FEDC constructs as well as the wild type promoter. Left: Predictions of the promoter activity by the mathematical model through variation of the β-catenin activity parameter B_{act} as well as TCDD concentration. Right: Corresponding measurements of the relative luciferase activity for increasing TCDD concentrations where iCRT3 modulated β-catenin activity.

the TCF/β-catenin binding site (Figure 3.22 middle right). Finally, when simulating the full model representing the wild type promoter, I observed a stronger 4.4-fold induction (Figure 3.22 bottom left) Here the inducibility limits were 35 % β-catenin activity and 2.5 nM TCDD. This closely resembled the results of the experiment

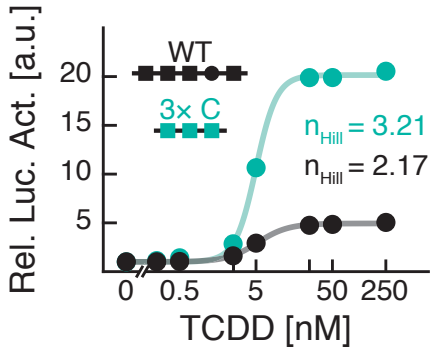


Figure 3.23: Human *CYP1A1* promoter was more sensitive to stimuli than synthetic promoters. Estimated Hill functions and coefficients for the data of the wild type promoter construct (black) as well as the synthetic triple C-DRE construct (green). Dots depict the relative luciferase activity for increasing TCDD concentrations while solid lines represent the Hill functions estimated from the data.

which displayed a maximal induction of 5 ± 0.44 relative luciferase activity (Figure 3.22 bottom right). For all three constructs I observed rather shallow stimulus-response curves in response to β-catenin and TCDD exposure that again resemble a logic AND-gate. Even though a complete β-catenin inhibition was experimentally not accessible, the data was in agreement with the predictions and likewise hints at a logic AND-gate integrating AhR and Wnt/β-catenin signaling. When we compared those measurements with the TCDD titration experiments, it was reassuring to see that these independent experiments were in good agreement (Figure A.3).

In summary, *CYP1A1* expression required the simultaneous presence of two signals, as a single stimulant — TCDD or β-catenin — was not sufficient to trigger a full response of the promoter. This behavior closely resembled a logic AND-gate and it was interesting to see that the apparent AND-gate logic of the *CYP1A1* promoter was established not only by interactions between the *cis*-regulatory binding sites, but in addition by the interaction between Wnt/β-catenin and AhR signaling pathways.

3.5.4 Sensitivity of the human *CYP1A1* promoter

Comparison of the experimental data and the predictions of the models from the natural promoter constructs with the synthetic promoter constructs displayed a less switch-like stimulus-response pattern in the natural promoter. Simpler to thermodynamic models a Hill function could provide information on the sensitivity of

the expression level (Kim et al., 2009). To quantify the differences in sigmoidality of the dose-response curves between the synthetic and the natural promoter constructs, I estimated their respective Hill coefficients. Hence, I fitted Hill functions to the dose-response curves of the synthetic 3x C-DRE as well as the natural wild type promoter construct (Figure 3.23). The resulting Hill coefficients were 3.21 and 2.17 for the synthetic and the natural construct, respectively. This result confirmed a more switch-like behavior of the synthetic constructs. Under consideration of the gradual AND-gate determined in the previous chapter, I concluded that human *CYP1A1* promoter responded in more sensitive manner to stimuli of the AhR and/or the Wnt/β-catenin signaling pathways.

3.5.5 Prediction of hepatic zonation

The previously observed reduced switch-like behavior of the natural promoter could be important for the zonated expression of *CYP1A1*. It was shown that the area of *CYP1A1* expression around the central vein is small and restricted to very high β-catenin activity if no AhR agonists are present (Bonin et al., 2006; Braeuning et al., 2007a; Hailfinger et al., 2006). In contrast, stimulation of AhR leads to an extension of this area into periportal regions. Similarly, the model displayed low expression of the reporter in cells that experienced 35 % β-catenin activity until TCDD concentration rose above 2.5 nM and further increased at higher doses (cf. Figure 3.22). On the other hand, cells with 100 % β-catenin activity exhibited a strong and saturating *CYP1A1* expression already at 5 nM TCDD.

To compare these predicted expression patterns with data, I mapped the predictions of the model for different TCDD concentrations on hexagonal grids that represented idealized liver lobules. For each of these idealized lobes I applied a linear gradient for β-catenin activity from the center (high activity) to the periphery (low activity) of each hexagon (Figure 3.24 top row). This type of gradient was deduced from immunohistochemical images showing zonation of active β-catenin in mouse liver (Benhamouche et al., 2006). For the TCDD concentration I assumed that TCDD will distribute evenly over the whole lobule. It should be noted that contrary to my assumption TCDD mainly injures the centrilobular regions (Gebhardt, 1992). But to increase comparability with 3-MC stimulation which mainly injures the periportal area (Gebhardt, 1992) a even TCDD distribution was assumed. My prediction resulted in no expression of *CYP1A1* if no TCDD was present. *CYP1A1*

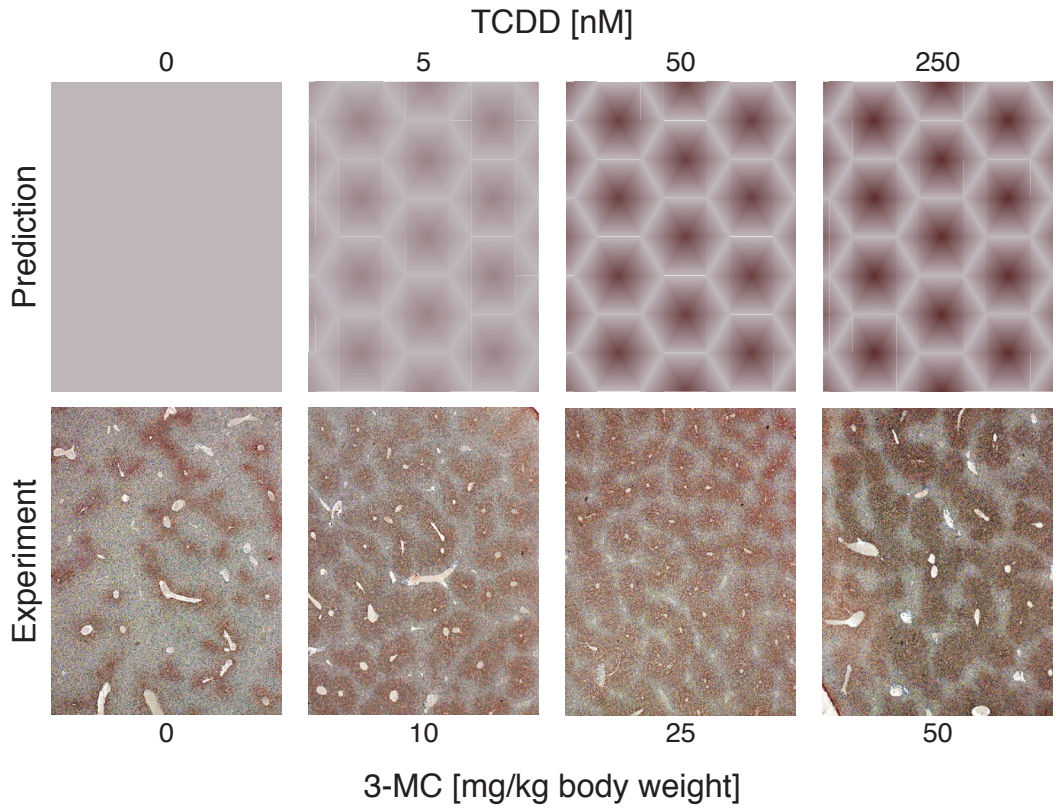


Figure 3.24: Model predicted spatial expression of *CYP1A1* as a result of its promoter logic. Top: Predictions of the mathematical model for the wild type construct for selected TCDD concentrations. Each hexagon represents an idealized hepatic lobule that was exposed to a β-catenin activity gradient from central vein (high) to portal vein (low). Bottom: Representative immunostainings of mouse liver for CYP1A for different concentrations of 3-MC. Darker browns correspond to increased expression.

was predicted to be slightly expressed in the pericentral zone for 5 nM TCDD. This area of expression widened for 50 nM TCDD whereas 250 nM TCDD only lead to increased expression but no additional extension of the expressing area in direction of the portal triad. We then performed immunostainings of CYP1A in mouse livers treated with increasing concentrations of the AhR inducer 3-MC (Figure 3.24 bottom row) to validate the predicted findings. When comparing these patterns I saw that the experimental data showed a basal perivenous expression of CYP1A for 0 mg/kg body weight 3-MC that was not predicted by the model. For 10 mg/kg body weight 3-MC a broadening of the CYP1A-expressing region around the central veins could be observed while higher concentrations of 3-MC only lead to an

increased expression intensity. This effect validated the predictions of the mathematical model.

In summary, the mathematical model was not only capable to quantitatively describe the behavior of the natural promoter in luciferase assays but had also qualitative predictive ability for the physiological expression patterns of *CYP1A1* observed in hepatic zonation.

4 Discussion

Intracorporal exposure to toxins such as TCDD trigger a wide variety of reactions. Biotransformation, mainly situated in the liver, renders xenobiotics less harmful and eases their excretion. Enzymes of the CYP superfamily such as CYP1A1 are the main drivers of this xenobiotic metabolism. Since regiospecific toxicity is observed for many xenobiotics (Gebhardt, 1992), their biotransforming enzymes are expressed in different zones as well. The pericentral expression of the *CYP1A1* gene exemplifies this zonation. Apart from drug metabolizing enzymes like CYP1A1, zonated activity is also observed for other proteins such as β-catenin. Along the portocentral axis activated β-catenin is mainly found around the central vein of liver lobules (Benhamouche et al., 2006). Braeuning and Schwarz (2010) found that a hepatocyte-specific knockout of β-catenin resulted in a loss of CYP1A expression in immunohistochemically stained mouse livers. Therefore I raised the question on which biological level this link between *CYP1A1* expression and β-catenin activity is implemented.

Modulation of *CYP1A1* expression by AhR and Wnt/β-catenin signaling

Recent studies presented evidence that β-catenin binds to the TCDD/AhR/Arnt complex (Braeuning et al., 2011; Kasai et al., 2013). With the help of co-immunoprecipitation we found that the interaction between AhR and β-catenin was increased upon TCDD treatment (Figure 3.3). This evidence further confirmed the binding of AhR and β-catenin and thus enables the Wnt/β-catenin signaling pathway to modulate *CYP1A1* expression as well. Additionally, Braeuning et al. (2011) found that within the *cis*-regulatory region of the *CYP1A1* promoter there exists a binding site that is targeted by the Wnt/β-catenin signaling pathway. Following from the structure of the promoter and the binding between AhR and β-catenin I asked how the TFs binding to the *cis*-regulatory region of the *CYP1A1* promoter cooperate, and how the two signaling pathways integrate into the promoter to achieve the

regulation of zonated gene expression.

Inconsistency of mutational impact on TF binding affinity

To analyze the *CYP1A1* promoter in more detail it was first necessary to understand how TFs associate with their specific target binding sites. With the help of ChIP-Seq data and publicly available databases I calculated the reduction in TF binding affinity when a mutation (like the ones we performed experimentally) was introduced. This bioinformatic analysis left me with an inconsistent picture of the mutational impact on TF binding affinity and thus omission of the therein generated results for the remaining work. While the affinity to the DREs is strongly reduced with data from Transfac and Jaspar it is only moderately reduced by using the ChIP-Seq dataset. This inconsistency could be due to the relatively short motifs stored in the Transfac and Jaspar databases. Gordân et al. (2013) for instance showed that the sequences flanking a well-defined core binding sequence can strongly effect binding affinity, mainly through influencing the 3D structure of the DNA binding site (Zhou et al., 2013). Since some proteins are able to bind DNA in different confirmations, Siggers and Gordân (2014) found that ChIP-Seq datasets are usually a mixture of different binding motifs which would result in unspecific binding models. For example, the NF-κB dimers binds to various low affinity sites that differ strongly from widely used binding models (Siggers et al., 2012; Wong et al., 2011). Another problem with the here used ChIP-Seq dataset was that the binding models were calculated by overlapping the binding events of AhR and Arnt alone (Lo and Matthews, 2012). Analysis of the heterodimeric TF complex as a whole could lead to fewer binding events, a more distinct binding motif and thus a more realistic representation of the binding affinity. The analysis of the TCF/β-catenin binding site on the other hand resulted in a consistent picture of binding affinity. Nevertheless, because of the lack of consistency for both, the DREs and the TCF/β-catenin binding site, I omitted the results of the bioinformatic analysis for the remaining study and assumed all-or-none binding processes between TFs and DNA. This assumption was also justified by other studies of TF-binding preferences. The comparison of protein-DNA crystal structures for wild type and mutant zinc finger protein Zif268/Egr1 for example showed that single amino acid mutations lead to altered protein side chain confirmations and thus, altered DNA-binding preferences (Miller and Pabo, 2001; Siggers and Gordân, 2014). Thus, not only binding site mutations

but also mutations in the amino acid sequence of TFs change their affinity to DNA. For future work it would be interesting to include data from high-throughput experiments, especially through the here shown calculation of the binding affinities. Furthermore the direct measurement of binding energies with protein-binding microarrays (Berger et al., 2006) would greatly reduce the complexity of the the modeling approaches and result in biologically more feasible models. These experiments however need to be carefully designed to overcome the pitfalls outlined above.

More-than-additive cooperativity between C- and D-DREs

As a starting point in answering the questions on TFBS cooperativity we point-mutated the binding sites sequences of the C- and D-DREs in the human *CYP1A1* promoter. I found that TCDD did not induce gene expression when the D-DRE alone was present but that it cooperated with the C-DRE to produce a more-than-additive induction of gene expression for the construct harboring both DREs (Figure 3.8). In the literature this cooperative behavior between TFBS's is a common motif (Casanovas et al., 2014; Giorgetti et al., 2010; Hermsen et al., 2010; Korenčič et al., 2012).

Short range cooperativity and sharp AND-gate-like signal integration of the synthetic promoters

To analyze the cooperativity between the C- and D-DREs in more detail we subsequently generated a library of synthetic promoter constructs. This library included nine promoter constructs with increasing amounts of the C- and D-DREs as well as four constructs with increasing distances between two C-DREs. Following transfection into mouse hepatoma cells and stimulation with increasing concentrations of TCDD we measured the luciferase activity. I found that the triple C-DRE construct was most strongly inducible with a decreasing inducibility for constructs holding four or more DREs (Figure 3.10). Furthermore, the D-DRE constructs responded only half as strong to TCDD than the C-DRE constructs. In concordance with results obtained from the bioinformatic analysis (Figure 3.9) I concluded that the D-DRE binding site underwent a weaker association with its TF than the C-DRE site.

In order to explain the observed behavior I devised a comprehensive two-part

mathematical model. The first part consisted of a simple description of the signaling events leading to the formation of the TCDD/AhR/Arnt/β-catenin TF complex (Figure 3.11). The second part represented the logic of the *cis*-regulatory region in the form of a thermodynamic model which related RNAP occupancy to gene expression (Figure 3.12). Within the thermodynamic model interactions between TFs themselves or between TFs and the RNAP were depicted with cooperative or adhesive binding energies, respectively. Following a Monte-Carlo approach for parameter estimation and profile likelihood estimation for determination of parameters identifiability, I used the resulting model parameters to reveal the hidden logic of the promoter. There I found that the TFs in the synthetic promoters only interacted with TFs that bind to nearby TFBS's (Figure 3.15). This short range interaction is commonly observed in *cis*-regulatory elements. Segal et al. (2008) for example utilized a thermodynamic model to successfully predict the expression patterns in the *Drosophila m.* embryo by assuming that cooperativity between two TFs decreases with distance. This phenomenon was also seen in studies of TF-DNA crystal structures. Garvie and Wolberger (2001) for example reported that the cooperativity between adjacent TFs stabilizes the interaction between the TF and DNA which enhances individual contributions to the transcriptional output. As a mechanism for this behavior, they identified alterations of residue-base contacts between protein and DNA. Kim et al. (2013) furthermore identified the deformation of the DNA structure as reason for the stabilization. In that respect I also saw that the induction of the constructs with an increasing distance between the C-DREs decreased validating the short range interaction observations (Figure 3.16). However, it should be noted that long range interactions are also observed in genomic promoters but almost exclusively in the form of DNA looping (Bintu et al., 2005). DNA looping was however not expected to happen in the experimental setup for the synthetic and natural promoter constructs.

Secondly, I found that only the TF binding to the DRE closest to the RNAP binding site underwent significant interaction with the RNAP while the remaining DREs did not or only insignificantly (Figure 3.15). Thus, I reasoned that the first DRE communicated the logic of the promoter, i.e., the binding configurations of the remaining DREs to the RNAP. This hypothesis was confirmed by the calculation of

the probability that the first DREs is occupied (Figure 3.17b). Through the calculation of the mean number of occupied binding sites in the synthetic promoters (Figure 3.17a), I furthermore found that the reduced induction of constructs containing four or more DREs could be explained by a sequestration effect.

The fully parameterized model was subsequently able to predict the integration of the AhR and the Wnt/β-catenin signaling pathways into the synthetic promoters. This prediction was successfully validated by experiments with a β-catenin inhibitor (Figure 3.18). I found that these two pathways were connected in the form of a sharp logic AND-gate. This boolean logic of signal integration is found in other promoters as well. Casanovas et al. (2014), for instance, showed that the hepcidin promoter with a mutated BRE2 binding site integrates the IL6 and BMP signaling as a logic AND-gate as well. But, this connection does not seem to have any physiological relevance since the BRE2 mutation is not (yet) implied with disease. For the classic *lacZYA* operon in *Escherichia coli*, Setty et al. (2003) discovered that the two inducers cAMP and IPTG are also connected by a logic AND-gate. They furthermore predicted that this AND-gate can be turned into a logic OR-gate through only a few mutations within the promoter. This prediction was later experimentally validated by their colleagues Mayo et al. (2006). In their groundbreaking work, Buchler et al. (2003) studied theoretically possible schemes of TF interactions that implement regulatory logic functions. It was for instance shown that two TFs that solely interact with the RNAP give rise to a boolean OR-gate. But, even more complex schemes such as XOR- or EQ-gates could be realized. Thus, knowing the relationship between promoter function and its architecture enables the design of programmable promoters for various tasks in synthetic biology (Cox et al., 2007). But in the context of the *CYP1A1* promoter there is currently no evidence in the literature that mutations in the DREs give rise to boolean logics other than AND-gates or even diseases.

Importance of the TCF/β-catenin TFBS and less switch-like AND-gate behavior of the natural promoters

In a next step I turned to the human *CYP1A1* promoter. Following generation of promoter constructs harboring different mutation patterns, transfection into 55.1c cells and stimulation with increasing concentrations of TCDD, we measured the luciferase activity. The resulting data exhibited a less obvious induction behavior as

compared to the synthetic promoters. Especially striking was the lower inducibility. Following the argumentation of the synthetic promoters this might be due to the larger distances of the TFBS's in the natural promoters. Nevertheless, the observed behavior could be quantitatively described with a comprehensive mathematical model (Figure 3.19). Only the inductions of the EDTC and FETC constructs were overestimated while the induction of the TC construct was underestimated. That may be the result of more complex cooperative interactions between TFs. While the thermodynamic modeling framework is limited to pairwise interactions between TFs the observed behavior might arise from interaction of three or more TFs. As of right now, there are no cases in the literature where this kind of interaction was included in the modeling approach. Additionally, the profile likelihood estimated one parameter to be practical non-identifiable (Figure 3.20). This non-identifiability could be resolved if this cooperative binding energy was measured directly. Another way to deal with the aforementioned shortcomings of the parameter estimation could be to expand the natural promoter library to reflect the complete combinatorial mutation pattern. In the present study, 18 reporter construct were used while there exist 2^5 possible mutational combinations.

From the resulting model parameters I found that as compared to the synthetic promoter the binding to RNAP became increasingly important in the natural promoters (Figure 3.21). Furthermore, I found that the binding site that is targeted by the Wnt/β-catenin signaling pathway was the most important cooperation partner for TFs associating with the DREs. After the physical interaction between AhR and β-catenin, this emphasized the second important mode of influence of the Wnt/β-catenin signaling pathway on *CYP1A1* expression. It was reassuring to see that interfering with the Wnt/β-catenin signaling pathway with glycogen synthase kinase inhibitors leads to similar results in *CYP1A1* expression (Briolotti et al., 2015).

When predicting the influence of β-catenin activity on *CYP1A1* expression I discovered that the two signaling pathways were again connected by a logic AND-gate (Figure 3.22). In the natural promoters the edges of the AND-gate were however less sharp, i.e., a more gradual change in expression occurred in comparison with the switch-like response of the synthetic promoters. The calculation of Hill coefficient furthermore validated this observation (Figure 3.23). This means that the human *CYP1A1* promoter is able to react in a more sensitive manner to changes in toxin exposure. Furthermore, changes β-catenin activity for example through the

presence of cancerous mutations can also be detected in a sensible way.

Prediction of the experimentally observed physiological zonation pattern

The observed sensitive response to toxin exposure and/or β-catenin titration was subsequently linked to the physiological zonation patterns observed in liver lobules. The model predicted gene expression patterns that were subsequently validated with immunostainings of mouse livers for CYP1A. Only, the basal perivenous expression of *CYP1A1* could not be captured by the model. Most likely, the reason for this was that AhR retains a basal activity in the liver due to the presence of low levels of endogenous AhR agonists. While the thermodynamic model can accurately describe the logic of the promoters in the luciferase assays, it should be noted that the fold-change of the *CYP1A1* gene is higher in the promoter of human cell lines (HepG2, CaCo-2, HT116) when compared to luciferase assays (Schreiber et al., 2006; Vaas et al., 2014). Thus, the model captured the essence of the promoter logic, but additional factors such as other binding sites or chromatin composition may influence the fold-change. It should also be noted that the kinetic details of the interactions in the promoter will determine how large the zone of expression will be and if the zone widens if toxins are present. While the described interaction of Wnt/β-catenin and AhR was sufficient to describe zonation of the *CYP1A1* expression, it is likely that the zonation of other genes is controlled by different and/or additional mechanisms. The expression of glutamine synthetase (GS), for example, is restricted to a small pericentral zone. This could be explained with an additional regulatory mechanism that silences an 5'-enhancer within the GS gene (Gaunitz et al., 2005).

Taken together, I reasoned that interactions between Wnt/β-catenin and AhR at the signaling level, together with the complex cooperative interactions between the DREs in the human *CYP1A1* promoter enable the spatiotemporal expression pattern observed *in vivo*.

5 Outlook

Even though the present study advanced the understanding of the *CYP1A1* promoter and the influences of two converging signaling pathways on hepatic zonation, there is still a lot of work to be done for a comprehensive understanding of this particular system as well as combinatorial gene expression as a whole. Since the present study was mainly carried out with *in vitro* luciferase reporter plasmids it would be interesting to investigate if the here observed mechanisms also hold in an *in vivo* setting. From an experimental point of view, this could be achieved by combining transgenic mouse models with an activated form of β-catenin (Braeuning, 2009) with genome editing techniques such as CRISPR/Cas-9 (Wang et al., 2013). With the so introduced mutations in the *cis*-regulatory regions of the *CYP1A1* promoter one could compare the expressional output with the results presented here. From a theoretical point of view it would be interesting to reverse the modeling approach. Here I based the layout of the model on prior biological knowledge. By combining a myriad of possible model structures with a model selection algorithm one could let the data dictate the composition of model which could spawn novel insights into the biology of the *CYP1A1* promoter. Furthermore, since nucleosomes play an important role gene expression (Kim and O'Shea, 2008; Lam et al., 2008), one would need to include their effect into a revamped formulation of the model as well.

Additionally, it would be interesting to integrate my model in larger models of liver zonation (Schliess et al., 2014) or even whole-body physiology based pharmacokinetic/pharmacodynamic models (Schwen et al., 2014; Thiel et al., 2015). For that it would be necessary to extend my model in space and time. Two studies in *Saccharomyces cerevisiae* and murine macrophages showed that (i) different TFs control the sequential propagation of the transcriptional program (Zenklusen et al., 2008), and (ii) thermodynamic models allow the study of dynamic gene regulation functions (Ramsey et al., 2008). The time dependency of a transcriptional

state could thus be modeled by combining the thermodynamic framework with rate equations (Dresch et al., 2013). The spatial component could be achieved by integration into a reaction-diffusion system. Consequently, it would help tremendously to experimentally quantify the exact concentrations of the TFs in the system.

And lastly, because of the strong ties between Wnt/β-catenin signaling and cancer (Polakis, 2000, 2007) the here presented work could aid the understanding and treatment of cancer. In particular, the optimal drug delivery and optimized activation/metabolization of drugs in hepatocellular carcinomas carrying a β-catenin mutation could be a promising application of my work.

A Appendix

A.1 Matrices of the thermodynamic models

In the following chapter the various matrices for the different models will be listed. With the help of Equation 2.25 the binding probabilities can be calculated. It should be noted that in the position matrices $\boldsymbol{L}_m$ and the state vectors $\boldsymbol{s}_m$ (with m representing the model identifiers) all occurring variables represent either TF concentrations ($[F]$ and $[T]$) or RNAP concentrations ($[P]$). Furthermore, all K_i are association constants to the specific binding site while the factors b_{ijk} collect the association constants as well as the statistic weight as introduced in Chapter 3.4.

Synthetic C-DRE promoters

The matrices for the synthetic promoter with only C-DREs present follow to:

$$
\boldsymbol{L}_C = \begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & [P] \\
1 & 1 & 1 & 1 & 1 & [F] & 1 \\
1 & 1 & 1 & 1 & 1 & [F] & [P] \\
1 & 1 & 1 & 1 & [F] & 1 & 1 \\
1 & 1 & 1 & 1 & [F] & 1 & [P] \\
1 & 1 & 1 & 1 & [F] & [F] & 1 \\
1 & 1 & 1 & 1 & [F] & [F] & [P] \\
1 & 1 & 1 & [F] & 1 & 1 & 1 \\
1 & 1 & 1 & [F] & 1 & 1 & [P] \\
1 & 1 & 1 & [F] & 1 & [F] & 1 \\
1 & 1 & 1 & [F] & 1 & [F] & [P] \\
1 & 1 & 1 & [F] & [F] & 1 & 1 \\
1 & 1 & 1 & [F] & [F] & 1 & [P] \\
1 & 1 & [F] & 1 & 1 & 1 & 1 \\
1 & 1 & [F] & 1 & 1 & 1 & [P] \\
1 & 1 & [F] & 1 & 1 & [F] & 1 \\
1 & 1 & [F] & 1 & 1 & [F] & [P] \\
1 & 1 & [F] & 1 & [F] & 1 & 1 \\
1 & 1 & [F] & 1 & [F] & 1 & [P] \\
1 & 1 & [F] & [F] & 1 & 1 & 1 \\
1 & 1 & [F] & [F] & 1 & 1 & [P] \\
1 & [F] & 1 & 1 & 1 & 1 & 1 \\
1 & [F] & 1 & 1 & 1 & 1 & [P] \\
1 & [F] & 1 & 1 & 1 & [F] & 1 \\
1 & [F] & 1 & 1 & 1 & [F] & [P] \\
1 & [F] & 1 & 1 & [F] & 1 & 1 \\
1 & [F] & 1 & 1 & [F] & 1 & [P] \\
1 & [F] & 1 & [F] & 1 & 1 & 1 \\
1 & [F] & 1 & [F] & 1 & 1 & [P] \\
1 & [F] & [F] & 1 & 1 & 1 & 1 \\
1 & [F] & [F] & 1 & 1 & 1 & [P] \\
[F] & 1 & 1 & 1 & 1 & 1 & 1 \\
[F] & 1 & 1 & 1 & 1 & 1 & [P] \\
[F] & 1 & 1 & 1 & 1 & [F] & 1 \\
[F] & 1 & 1 & 1 & 1 & [F] & [P] \\
[F] & 1 & 1 & 1 & [F] & 1 & 1 \\
[F] & 1 & 1 & 1 & [F] & 1 & [P] \\
[F] & 1 & 1 & [F] & 1 & 1 & 1 \\
[F] & 1 & 1 & [F] & 1 & 1 & [P] \\
[F] & 1 & [F] & 1 & 1 & 1 & 1 \\
[F] & 1 & [F] & 1 & 1 & 1 & [P] \\
[F] & [F] & 1 & 1 & 1 & 1 & 1 \\
[F] & [F] & 1 & 1 & 1 & 1 & [P]
\end{bmatrix}, \quad
\boldsymbol{t}_C = \begin{bmatrix}
0 \\ 1 \\ 0 \\ 1
\end{bmatrix}, \quad
\boldsymbol{s}_C = \begin{bmatrix}
1 \\ [P] \\ [F] \\ [F]\,[P] \\ [F] \\ [F]\,[P] \\ [F]^2 \\ [F]^2\,[P] \\ [F] \\ [F]\,[P] \\ [F]^2 \\ [F]^2\,[P] \\ [F]^2 \\ [F]^2\,[P] \\ [F] \\ [F]\,[P] \\ [F]^2 \\ [F]^2\,[P] \\ [F]^2 \\ [F]^2\,[P] \\ [F]^2 \\ [F]^2\,[P] \\ [F] \\ [F]\,[P] \\ [F]^2 \\ [F]^2\,[P] \\ [F]^2 \\ [F]^2\,[P] \\ [F]^2 \\ [F]^2\,[P] \\ [F]^2 \\ [F]^2\,[P] \\ [F] \\ [F]\,[P] \\ [F]^2 \\ [F]^2\,[P] \\ [F]^2 \\ [F]^2\,[P] \\ [F]^2 \\ [F]^2\,[P] \\ [F]^2 \\ [F]^2\,[P] \\ [F]^2 \\ [F]^2\,[P]
\end{bmatrix}, \quad
\boldsymbol{b}_C = \begin{bmatrix}
1 \\ K_P \\ K_C \\ b_{PC_1} \\ K_C \\ b_{PC_2} \\ b_{C_{d=1}} \\ b_{PC_1C_2} \\ K_C \\ b_{PC_3} \\ b_{C_{d=2}} \\ b_{PC_1C_3} \\ b_{C_{d=1}} \\ b_{PC_2C_3} \\ K_C \\ b_{PC_4} \\ b_{C_{d=3}} \\ b_{PC_1C_4} \\ b_{C_{d=2}} \\ b_{PC_2C_4} \\ b_{C_{d=1}} \\ b_{PC_3C_4} \\ K_C \\ b_{PC_5} \\ b_{C_{d=4}} \\ b_{PC_1C_5} \\ b_{C_{d=3}} \\ b_{PC_2C_5} \\ b_{C_{d=2}} \\ b_{PC_3C_5} \\ b_{C_{d=1}} \\ b_{PC_4C_5} \\ K_C \\ b_{PC_6} \\ b_{C_{d=5}} \\ b_{PC_1C_6} \\ b_{C_{d=4}} \\ b_{PC_2C_6} \\ b_{C_{d=3}} \\ b_{PC_3C_6} \\ b_{C_{d=2}} \\ b_{PC_4C_6} \\ b_{C_{d=1}} \\ b_{PC_5C_6}
\end{bmatrix}.
$$

The columns in the position matrix $\boldsymbol{L}_C$ describe the TFBS's 6 to 1 as well as the RNAP binding site on the reporter construct.

Synthetic D-DRE promoters

The matrices for the synthetic promoter with only D-DREs present follow to:

$$\boldsymbol{L}_D = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & [P] \\ 1 & 1 & [F] & 1 \\ 1 & 1 & [F] & [P] \\ 1 & [F] & 1 & 1 \\ 1 & [F] & 1 & [P] \\ 1 & [F] & [F] & 1 \\ 1 & [F] & [F] & [P] \\ [F] & 1 & 1 & 1 \\ [F] & 1 & 1 & [P] \\ [F] & 1 & [F] & 1 \\ [F] & 1 & [F] & [P] \\ [F] & [F] & 1 & 1 \\ [F] & [F] & 1 & [P] \end{bmatrix}, \boldsymbol{t}_D = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \boldsymbol{s}_D = \begin{bmatrix} 1 \\ [P] \\ [F] \\ [F]\,[P] \\ [F] \\ [F]\,[P] \\ [F]^2 \\ [F]^2\,[P] \\ [F] \\ [F]\,[P] \\ [F]^2 \\ [F]^2\,[P] \\ [F]^2 \\ [F]^2\,[P] \end{bmatrix}, \boldsymbol{b}_D = \begin{bmatrix} 1 \\ K_P \\ K_D \\ b_{PD_1} \\ K_D \\ b_{PD_2} \\ b_{D_{d=1}} \\ b_{PD_1D_2} \\ K_D \\ b_{PD_3} \\ b_{D_{d=2}} \\ b_{PD_1D_3} \\ b_{D_{d=1}} \\ b_{PD_2D_3} \end{bmatrix}$$

The columns in the position matrix $\boldsymbol{L}_D$ describe the TFBS's 3 to 1 as well as the RNAP binding site on the reporter construct.

Natural promoter

The matrices for the natural promoter are:

$$\boldsymbol{L}_N = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & [P] \\ 1 & 1 & 1 & 1 & [F] & 1 \\ 1 & 1 & 1 & 1 & [F] & [P] \\ 1 & 1 & 1 & [T] & 1 & 1 \\ 1 & 1 & 1 & [T] & 1 & [P] \\ 1 & 1 & 1 & [T] & [F] & 1 \\ 1 & 1 & 1 & [T] & [F] & [P] \\ 1 & 1 & [F] & 1 & 1 & 1 \\ 1 & 1 & [F] & 1 & 1 & [P] \\ 1 & 1 & [F] & 1 & [F] & 1 \\ 1 & 1 & [F] & 1 & [F] & [P] \\ 1 & 1 & [F] & [T] & 1 & 1 \\ 1 & 1 & [F] & [T] & 1 & [P] \\ 1 & [F] & 1 & 1 & 1 & 1 \\ 1 & [F] & 1 & 1 & 1 & [P] \\ 1 & [F] & 1 & 1 & [F] & 1 \\ 1 & [F] & 1 & 1 & [F] & [P] \\ 1 & [F] & 1 & [T] & 1 & 1 \\ 1 & [F] & 1 & [T] & 1 & [P] \\ 1 & [F] & [F] & 1 & 1 & 1 \\ 1 & [F] & [F] & 1 & 1 & [P] \\ [F] & 1 & 1 & 1 & 1 & 1 \\ [F] & 1 & 1 & 1 & 1 & [P] \\ [F] & 1 & 1 & 1 & [F] & 1 \\ [F] & 1 & 1 & 1 & [F] & [P] \\ [F] & 1 & 1 & [T] & 1 & 1 \\ [F] & 1 & 1 & [T] & 1 & [P] \\ [F] & 1 & [F] & 1 & 1 & 1 \\ [F] & 1 & [F] & 1 & 1 & [P] \\ [F] & [F] & 1 & 1 & 1 & 1 \\ [F] & [F] & 1 & 1 & 1 & [P] \end{bmatrix}, \boldsymbol{t}_N = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \boldsymbol{s}_N = \begin{bmatrix} 1 \\ [P] \\ [F] \\ [F]\,[P] \\ [T] \\ [T]\,[P] \\ [T]\,[F] \\ [T]\,[F]\,[P] \\ [F] \\ [F]\,[P] \\ [F]^2 \\ [F]^2\,[P] \\ [F]\,[T] \\ [F]\,[T]\,[P] \\ [F] \\ [F]\,[P] \\ [F]^2 \\ [F]^2\,[P] \\ [F]\,[T] \\ [F]\,[T]\,[P] \\ [F]^2 \\ [F]^2\,[P] \\ [F] \\ [F]\,[P] \\ [F]^2 \\ [F]^2\,[P] \\ [F]\,[T] \\ [F]\,[T]\,[P] \\ [F]^2 \\ [F]^2\,[P] \\ [F]^2 \\ [F]^2\,[P] \end{bmatrix}, \boldsymbol{b}_N = \begin{bmatrix} 1 \\ K_P \\ K_C \\ b_{PC} \\ K_T \\ b_{PT} \\ b_{CT} \\ b_{PCT} \\ K_D \\ b_{PD} \\ b_{CD} \\ b_{PCD} \\ b_{DT} \\ b_{PDT} \\ K_E \\ b_{PE} \\ b_{CE} \\ b_{PCE} \\ b_{TE} \\ b_{PTE} \\ b_{DE} \\ b_{PDE} \\ K_T \\ b_{PF} \\ b_{CF} \\ b_{PCF} \\ b_{TF} \\ b_{PTF} \\ b_{DF} \\ b_{PDF} \\ b_{EF} \\ b_{PEF} \end{bmatrix}$$

The columns in the position matrix $\boldsymbol{L}_N$ describe the DREs F to D, the TCF/β-catenin binding site, the C-DRE as well as the RNAP binding site on the reporter construct.

A.2 Supplementary Tables

Estimated parameters of the synthetic (Table A.1) as well as the natural (Table A.2) models.

Parameter		Confidence intervals		
Name	Value	σ^-	σ^+	Unit
B_{Tot}	17.3884	0.4184	0.3912	M
K_2	70.5033	36.6617	45.1221	M
α	71.4501	22.8640	57.1601	-
q_P	0.0204	0.0006	0.0006	a.u.
K_C	0.0012	$3.3703 \cdot 10^{-05}$	$3.2503 \cdot 10^{-05}$	-
K_D	0.0003	$1.2495 \cdot 10^{-05}$	$1.2029 \cdot 10^{-05}$	-
ϵ_{1P}	-3.4146	0.0173	0.0181	k_BT
ϵ_{2P}	0	0.0746	∞	k_BT
ϵ_{3P}	-0.5487	0.0198	0.0207	k_BT
ϵ_{4P}	0	0.9083	∞	k_BT
ϵ_{5P}	0	1.0578	∞	k_BT
ϵ_{6P}	0	2.1972	∞	k_BT
$\epsilon_{d=1}$	-7.7567	0.0862	0.1012	k_BT
$\epsilon_{d=2}$	-9.2748	0.0392	0.0433	k_BT
$\epsilon_{d=3}$	0	6.9907	∞	k_BT
$\epsilon_{d=4}$	0	7.0277	∞	k_BT
$\epsilon_{d=5}$	0	8.1437	∞	k_BT

Table A.1: Parameter set of the synthetic constructs with point-wise confidence intervals as estimated from profile likelihood.

Parameter		Confidence intervals		
Name	Value	σ^-	σ^+	Unit
B_{Tot}	0.5353	0.0095	0.0087	M
K_2	3.0355	0.2550	0.2793	M
α	0.3200	0.0279	0.0272	-
q_P	0	∞	0.0039	a.u.
K_T	$2.8032 \cdot 10^{-12}$	$1.3446 \cdot 10^{-13}$	$1.2614 \cdot 10^{-13}$	-
K_E	$1.0633 \cdot 10^{-06}$	$5.8648 \cdot 10^{-08}$	$5.5825 \cdot 10^{-08}$	-
K_F	0.3135	0.0502	0.0564	-
ϵ_{CP}	-7.0610	0.0416	0.0477	k_BT
ϵ_{TP}	-14.0765	0.0440	0.0492	k_BT
ϵ_{DP}	0	0.0630	∞	k_BT
ϵ_{EP}	-12.8453	0.0512	0.0567	k_BT
ϵ_{FP}	-1.0061	0.0606	0.0685	k_BT
ϵ_{TC}	-13.8069	0.0746	0.0851	k_BT
ϵ_{DC}	-8.1055	0.3075	0.4463	k_BT
ϵ_{DT}	-20.7887	0.2469	0.3285	k_BT
$\epsilon_{EC}*$	0	0	0	k_BT
ϵ_{ET}	-15.2893	0.0746	0.0854	k_BT
ϵ_{ED}	-8.7341	0.5878	∞	k_BT
$\epsilon_{FC}*$	0	0	0	k_BT
$\epsilon_{FT}*$	0	0	0	k_BT
ϵ_{FD}	-10.0960	0.3365	0.3567	k_BT
ϵ_{FE}	-2.5828	0.1570	0.1863	k_BT

Table A.2: Parameter set of the natural promoter constructs with point-wise confidence intervals as estimated from profile likelihood. The parameters marked with an asterisk were set to zero because they qualified as long-distance interactions.

A.3 Supplementary Figures

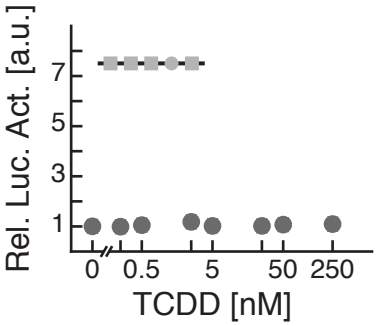


Figure A.1: The reporter construct with no wild type binding sites was not inducible by TCDD. Relative luciferase activity of the empty reporter construct for increasing TCDD concentrations is shown (gray dots).

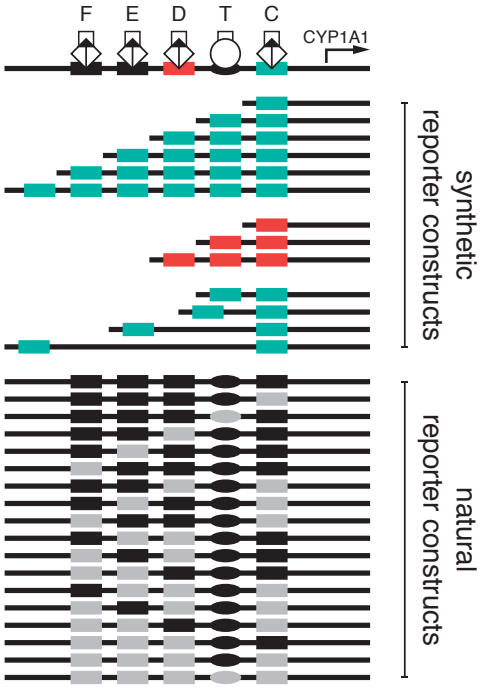


Figure A.2: Synthetic and natural reporter constructs. 13 synthetic reporter constructs harboring 1-6 C-DRE, 1-3 D-DREs and the distances 19 bp, 49 bp, 156 bp and 292 bp between two C-DREs. 18 natural reporter constructs with different mutation patterns. Black ellipses and rectangles represent wild type binding sites while mutated TFBS's are displayed in gray.

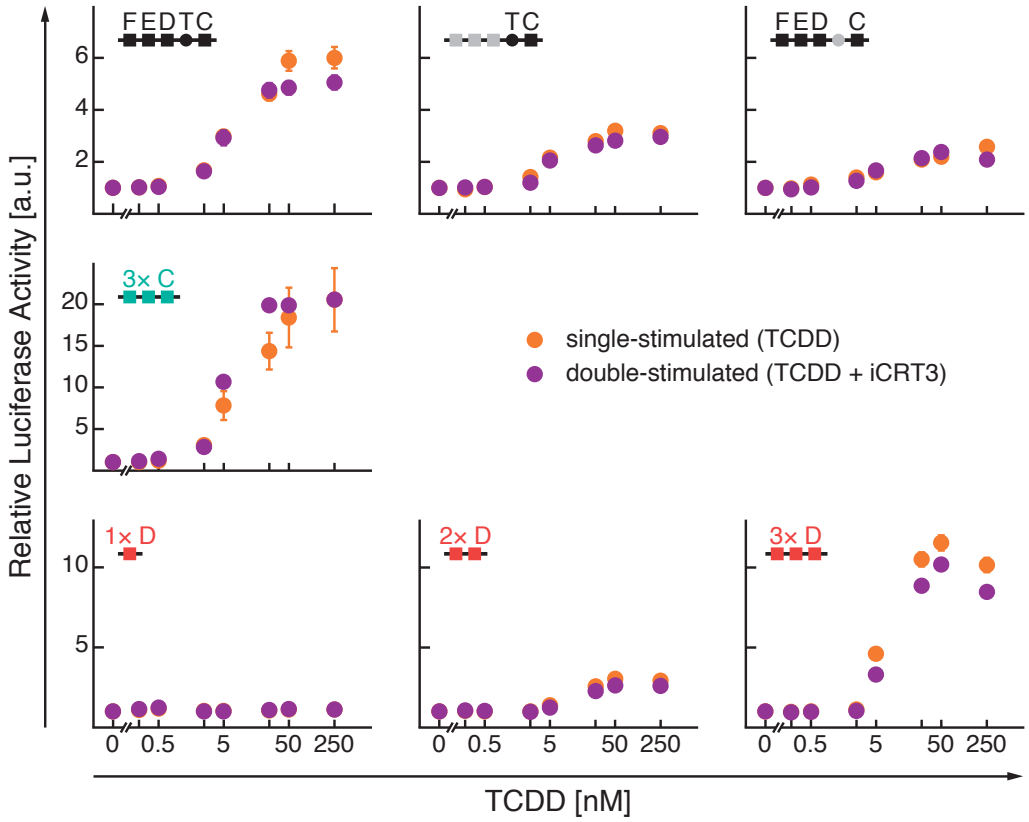


Figure A.3: Single and double-stimulated datasets were consistent. The TCDD concentration series of three natural (wild type, TC and FEDC) and four synthetic (3x C-DRE and 1x-3x D-DRE) constructs were extracted from the corresponding datasets. Single-stimulated data (orange) was taken from Figures 3.10 and 3.19 while double-stimulated data (prurple) was tabken from the 100 % β-catenin activity rows in Figures 3.18 and 3.22.

Bibliography

Aberle, H., Bauer, A., Stappert, J., Kispert, A., and Kemler, R. β-catenin is a target for the ubiquitin-proteasome pathway. *The EMBO journal*, 16(13):3797–3804, July 1997.

Ackers, G. K., Johnson, A. D., and Shea, M. A. Quantitative model for gene regulation by lambda phage repressor. *Proceedings of the National Academy of Sciences of the United States of America*, 79(4):1129–1133, Feb. 1982.

Amit, I., Citri, A., Shay, T., Lu, Y., Katz, M., Zhang, F., Tarcic, G., Siwak, D., Lahad, J., Jacob-Hirsch, J., Amariglio, N., Vaisman, N., Segal, E., Rechavi, G., Alon, U., Mills, G. B., Domany, E., and Yarden, Y. A module of negative feedback regulators defines growth factor signaling. *Nature genetics*, 39(4):503–512, Apr. 2007.

Anastas, J. N. and Moon, R. T. WNT signalling pathways as therapeutic targets in cancer. *Nature reviews. Cancer*, 13(1):11–26, Jan. 2013.

Antonsson, C., Whitelaw, M. L., McGuire, J., Gustafsson, J. A., and Poellinger, L. Distinct roles of the molecular chaperone hsp90 in modulating dioxin receptor function via the basic helix-loop-helix and PAS domains. *Molecular and cellular biology*, 15(2):756–765, Feb. 1995.

Bacsi, S. G., Reisz-Porszasz, S., and Hankinson, O. Orientation of the heterodimeric aryl hydrocarbon (dioxin) receptor complex on its asymmetric DNA recognition sequence. *Molecular pharmacology*, 47(3):432–438, Mar. 1995.

Baron, R. and Kneissel, M. WNT signaling in bone homeostasis and disease: from human mutations to treatments. *Nature Medicine*, 19(2):179–192, Feb. 2013.

Benhamouche, S., Decaens, T., Godard, C., Chambrey, R., Rickman, D. S., Moinard, C., Vasseur-Cognet, M., Kuo, C. J., Kahn, A., Perret, C., and Colnot, S. Apc tumor

suppressor gene is the "zonation-keeper" of mouse liver. *Developmental Cell*, 10 (6):759–770, June 2006.

Benos, P. V., Bulyk, M. L., and Stormo, G. D. Additivity in protein-DNA interactions: how good an approximation is it? *Nucleic Acids Research*, 30(20):4442–4451, Oct. 2002.

Berg, O. G. and Hippel, P. H. Selection of DNA binding sites by regulatory proteins. Statistical-mechanical theory and application to operators and promoters. *Journal of molecular biology*, 193(4):723–750, Feb. 1987.

Berger, M. F., Philippakis, A. A., Qureshi, A. M., He, F. S., Estep, P. W., and Bulyk, M. L. Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nature biotechnology*, 24(11):1429–1435, Nov. 2006.

Bhanot, P., Brink, M., Samos, C. H., Hsieh, J.-C., Wang, Y., Macke, J. P., Andrew, D., Nathans, J., and Nusse, R. A new member of the frizzled family from Drosophila functions as a Wingless receptor. *Nature*, 382(6588):225–230, July 1996.

Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., and Phillips, R. Transcriptional regulation by the numbers: models. *Current opinion in genetics & development*, 15(2):116–124, Dec. 2004.

Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., Kuhlman, T., and Phillips, R. Transcriptional regulation by the numbers: applications. *Current opinion in genetics & development*, 15(2):125–135, Apr. 2005.

Bock, K. W. Aryl hydrocarbon or dioxin receptor: biologic and toxic responses. *Reviews of physiology, biochemistry and pharmacology*, 125:1–42, 1994.

Bonin, M., Braeuning, A., Ittrich, C., Köhle, C., Hailfinger, S., Buchmann, A., and Schwarz, M. Differential gene expression in periportal and perivenous mouse hepatocytes. *The FEBS journal*, 273(22):5051–5061, Nov. 2006.

Bosisio, D., Marazzi, I., Agresti, A., Shimizu, N., Bianchi, M. E., and Natoli, G. A hyper-dynamic equilibrium between promoter-bound and nucleoplasmic dimers controls NF-κB-dependent gene activity. *The EMBO journal*, 25(4):798–810, Feb. 2006.

Bourhis, E., Tam, C., Franke, Y., Bazan, J. F., Ernst, J., Hwang, J., Costa, M., Cochran, A. G., and Hannoush, R. N. Reconstitution of a Frizzled8-Wnt3a-LRP6 signaling complex reveals multiple Wnt and Dkk1 binding sites on LRP6. *The Journal of biological chemistry*, 285(12):9172–9179, Mar. 2010.

Braeuning, A. Regulation of cytochrome P450 expression by Ras- and beta-catenin-dependent signaling. *Current drug metabolism*, 10(2):138–158, Feb. 2009.

Braeuning, A. and Buchmann, A. The glycogen synthase kinase inhibitor 3-(2,4-dichlorophenyl)-4-(1-methyl-1H-indol-3-yl)-1H-pyrrole-2,5-dione (SB216763) is a partial agonist of the aryl hydrocarbon receptor. *Drug metabolism and disposition: the biological fate of chemicals*, 37(8):1576–1580, Aug. 2009.

Braeuning, A. and Schwarz, M. β-catenin as a multilayer modulator of zonal cytochrome P450 expression in mouse liver. *Biological chemistry*, 391(2-3):139–148, Feb. 2010.

Braeuning, A. and Vetter, S. The nuclear factor κB inhibitor (E)-2-fluoro-4'-methoxystilbene inhibits firefly luciferase. *Bioscience reports*, 32(6):531–537, Dec. 2012.

Braeuning, A., Ittrich, C., Köhle, C., Buchmann, A., and Schwarz, M. Zonal gene expression in mouse liver resembles expression patterns of Ha-ras and beta-catenin mutated hepatomas. *Drug metabolism and disposition: the biological fate of chemicals*, 35(4):503–507, Apr. 2007a.

Braeuning, A., Menzel, M., Kleinschnitz, E.-M., Harada, N., Tamai, Y., Köhle, C., Buchmann, A., and Schwarz, M. Serum components and activated Ha-ras antagonize expression of perivenous marker genes stimulated by β-catenin signaling in mouse hepatocytes. *The FEBS journal*, 274(18):4766–4777, Sept. 2007b.

Braeuning, A., Sanna, R., Huelsken, J., and Schwarz, M. Inducibility of drug-metabolizing enzymes by xenobiotics in mice with liver-specific knockout of Ctnnb1. *Drug metabolism and disposition: the biological fate of chemicals*, 37(5):1138–1145, May 2009.

Braeuning, A., Singh, Y., Rignall, B., Buchmann, A., Hammad, S., Othman, A., von Recklinghausen, I., Godoy, P., Hoehme, S., Drasdo, D., Hengstler, J. G., and

Schwarz, M. Phenotype and growth behavior of residual β-catenin-positive hepatocytes in livers of β-catenin-deficient mice. *Histochemistry and cell biology*, 134 (5):469–481, Nov. 2010.

Braeuning, A., Köhle, C., Buchmann, A., and Schwarz, M. Coordinate regulation of cytochrome P450 1a1 expression in mouse liver by the aryl hydrocarbon receptor and the β-catenin pathway. *Toxicological sciences*, 122(1):16–25, July 2011.

Braeuning, A., Vetter, S., Orsetti, S., and Schwarz, M. Paradoxical cytotoxicity of tert-butylhydroquinone in vitro: What kills the untreated cells? *Archives of toxicology*, 86(9):1481–1487, Sept. 2012.

Briolotti, P., Chaloin, L., Balaguer, P., Da Silva, F., Tománková, V., Pascussi, J.-M., Duret, C., Fabre, J.-M., Ramos, J., Klieber, S., Maurel, P., Daujat-Chavanieu, M., and Gerbal-Chaloin, S. Analysis of glycogen synthase kinase inhibitors that regulate cytochrome P450 expression in primary human hepatocytes by activation of β-catenin, aryl hydrocarbon receptor and pregnane X receptor signaling. *Toxicological Sciences*, page kfv177, Aug. 2015.

Brown, C. D., Johnson, D. S., and Sidow, A. Functional architecture and evolution of transcriptional elements that drive gene coexpression. *Science*, 317(5844):1557–1560, Sept. 2007.

Buchler, N. E., Gerland, U., and Hwa, T. On schemes of combinatorial transcription logic. *Proceedings of the National Academy of Sciences of the United States of America*, 100(9):5136–5141, Apr. 2003.

Bussemaker, H. J., Foat, B. C., and Ward, L. D. Predictive modeling of genome-wide mRNA expression: from modules to molecules. *Annual review of biophysics and biomolecular structure*, 36:329–347, 2007.

Buters, J. T. M. Phase I Metabolism. In Snyder, R. and Greim, H., editors, *Toxicology and Risk Assessment*. Wiley-Interscience, Chichester, UK, Jan. 2008.

Cadigan, K. M. and Peifer, M. Wnt signaling from development to disease: insights from model systems. *Cold Spring Harbor perspectives in biology*, 1(2):a002881, Aug. 2009.

Cadigan, K. M. and Waterman, M. L. TCF/LEFs and Wnt signaling in the nucleus. *Cold Spring Harbor perspectives in biology*, 4(11):a007906–a007906, Nov. 2012.

Campbell, C. T. and Kim, G. SPR microscopy and its applications to high-throughput analyses of biomolecular binding events and their kinetics. *Biomaterials*, 28(15):2380–2392, May 2007.

Carver, L. A. and Bradfield, C. A. Ligand-dependent interaction of the aryl hydrocarbon receptor with a novel immunophilin homolog in vivo. *The Journal of biological chemistry*, 272(17):11452–11456, Apr. 1997.

Cary, M. P., Bader, G. D., and Sander, C. Pathway information for systems biology. *FEBS Letters*, 579(8):1815–1820, Mar. 2005.

Casanovas, G., Banerji, A., d'Alessio, F., Muckenthaler, M. U., and Legewie, S. A multi-scale model of hepcidin promoter regulation reveals factors controlling systemic iron homeostasis. *PLoS computational biology*, 10(1):e1003421, Jan. 2014.

Cavallo, R. A., Cox, R. T., Moline, M. M., Roose, J., Polevoy, G. A., Peifer, M., and Bejsovec, A. Drosophila Tcf and Groucho interact to repress Wingless signalling activity. *Nature*, 395(6702):604–608, Oct. 1998.

Chen, Q. K., Hertz, G. Z., and Stormo, G. D. MATRIX SEARCH 1.0: a computer program that scans DNA sequences for transcriptional elements using a database of weight matrices. *Computer applications in the biosciences : CABIOS*, 11(5):563–566, Oct. 1995.

Chen, S., Bubeck, D., MacDonald, B. T., Liang, W.-X., Mao, J.-H., Malinauskas, T., Llorca, O., Aricescu, A. R., Siebold, C., He, X., and Jones, E. Y. Structural and Functional Studies of LRP6 Ectodomain Reveal a Platform for Wnt Signaling. *Developmental Cell*, 21(5):848–861, Nov. 2011.

Cheng, C., Alexander, R., Min, R., Leng, J., Yip, K. Y., Rozowsky, J., Yan, K.-K., Dong, X., Djebali, S., Ruan, Y., Davis, C. A., Carninci, P., Lassman, T., Gingeras, T. R., Guigó, R., Birney, E., Weng, Z., Snyder, M., and Gerstein, M. Understanding transcriptional regulation by integrative analysis of transcription factor binding data. *Genome research*, 22(9):1658–1667, Sept. 2012.

Chesire, D. R., Dunn, T. A., Ewing, C. M., Luo, J., and Isaacs, W. B. Identification of aryl hydrocarbon receptor as a putative Wnt/β-catenin pathway target gene in prostate cancer cells. *Cancer Research*, 64(7):2523–2533, Apr. 2004.

Clevers, H. Wnt/β-catenin signaling in development and disease. *Cell*, 127(3):469–480, Nov. 2006.

Clevers, H. and Nusse, R. Wnt/β-catenin signaling and disease. *Cell*, 149(6):1192–1205, June 2012.

Cobelli, C. and DiStefano, J. J. Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. *The American journal of physiology*, 239 (1):R7–24, July 1980.

Colletti, M., Cicchini, C., Conigliaro, A., Santangelo, L., Alonzi, T., Pasquini, E., Tripodi, M., and Amicone, L. Convergence of Wnt signaling on the HNF4alpha-driven transcription in controlling liver zonation. *Gastroenterology*, 137(2):660–672, Aug. 2009.

Corchero, J., Pimprale, S., Kimura, S., and Gonzalez, F. J. Organization of the CYP1A cluster on human chromosome 15: implications for gene regulation. *Pharmacogenetics*, 11(1):1–6, Feb. 2001.

Cox, R. S., Surette, M. G., and Elowitz, M. B. Programming gene expression with combinatorial promoters. *Molecular systems biology*, 3:145, 2007.

Dann, C. E., Hsieh, J.-C., Rattner, A., Sharma, D., Nathans, J., and Leahy, D. J. Insights into Wnt binding and signalling from the structures of two Frizzled cysteine-rich domains. *Nature*, 412(6842):86–90, July 2001.

Darzacq, X., Shav-Tal, Y., de Turris, V., Brody, Y., Shenoy, S. M., Phair, R. D., and Singer, R. H. In vivo dynamics of RNA polymerase II transcription. *Nature Structural and Molecular Biology*, 14(9):796–806, Sept. 2007.

Davarinos, N. A. and Pollenz, R. S. Aryl hydrocarbon receptor imported into the nucleus following ligand binding is rapidly degraded via the cytosplasmic proteasome following nuclear export. *The Journal of biological chemistry*, 274(40): 28708–28715, Oct. 1999.

de Jong, H. Modeling and Simulation of Genetic Regulatory Systems: A Literature Review. *Journal of computational biology*, 9(1):67–103, Jan. 2002.

Denis, M., Cuthill, S., Wikström, A. C., Poellinger, L., and Gustafsson, J. A. Association of the dioxin receptor with the Mr 90,000 heat shock protein: a structural kinship with the glucocorticoid receptor. *Biochemical and Biophysical Research Communications*, 155(2):801–807, Sept. 1988.

D'haeseleer, P. What are DNA sequence motifs? *Nature biotechnology*, 24(4):423–425, Apr. 2006.

Dill, K. and Bromberg, S. *Molecular Driving Forces*. Statistical Thermodynamics in Chemistry and Biology. Garland Science, 2003.

Dresch, J. M., Thompson, M. A., Arnosti, D. N., and Chiu, C. Two-Layer Mathematical Modeling of Gene Expression: Incorporating DNA-Level Information and System Dynamics. *SIAM Journal on Applied Mathematics*, 73(2):804–826, Apr. 2013.

Durbin, R., Eddy, S. R., Krogh, A., and Mitchison, G. *Biological Sequence Analysis*. Probabilistic Models of Proteins and Nucleic Acids. Cambridge University Press, Apr. 1998.

Ellis, T., Wang, X., and Collins, J. J. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nature biotechnology*, 27(5): 465–471, May 2009.

ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414):57–74, Sept. 2012.

Foat, B. C., Morozov, A. V., and Bussemaker, H. J. Statistical mechanical modeling of genome-wide transcription factor occupancy data by MatrixREDUCE. *Bioinformatics*, 22(14):e141–9, July 2006.

Frevert, U., Engelmann, S., Zougbédé, S., Stange, J., Ng, B., Matuschewski, K., Liebes, L., and Yee, H. Intravital observation of Plasmodium berghei sporozoite infection of the liver. *PLoS Biology*, 3(6):e192, June 2005.

Garcia, H. G., Kondev, J., Orme, N., Theriot, J. A., and Phillips, R. Thermodynamics of biological processes. *Methods in enzymology*, 492:27–59, Dec. 2010a.

Garcia, H. G., Sanchez, A. A., Kuhlman, T. T., Kondev, J. J., and Phillips, R. R. Transcription by the numbers redux: experiments and calculations that surprise. *Trends in cell biology*, 20(12):11–11, Nov. 2010b.

Garvie, C. W. and Wolberger, C. Recognition of specific DNA sequences. *Molecular Cell*, 8(5):937–946, Nov. 2001.

Gaunitz, F., Deichsel, D., Heise, K., Werth, M., Anderegg, U., and Gebhardt, R. An intronic silencer element is responsible for specific zonal expression of glutamine synthetase in the rat liver. *Hepatology*, 41(6):1225–1232, June 2005.

Gebhardt, R. Metabolic zonation of the liver: regulation and implications for liver function. *Pharmacology & therapeutics*, 53(3):275–354, 1992.

Gebhardt, R. and Matz-Soja, M. Liver zonation: Novel aspects of its regulation and its impact on homeostasis. *World Journal of Gastroenterology : WJG*, 20(26): 8491–8504, July 2014.

Gebhardt, R., Baldysiak-Figiel, A., Krügel, V., Ueberham, E., and Gaunitz, F. Hepatocellular expression of glutamine synthetase: an indicator of morphogen actions as master regulators of zonation in adult liver. *Progress in histochemistry and cytochemistry*, 41(4):201–266, 2007.

Gertz, J. and Cohen, B. A. Environment-specific combinatorial cis-regulation in synthetic promoters. *Molecular systems biology*, 5(1):244, 2009.

Gertz, J., Siggia, E. D., and Cohen, B. A. Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature*, 457(7226):215–218, Nov. 2008.

Giorgetti, L., Siggers, T., Tiana, G., Caprara, G., Notarbartolo, S., Corona, T., Pasparakis, M., Milani, P., Bulyk, M. L., and Natoli, G. Noncooperative interactions between transcription factors and clustered DNA binding sites enable graded transcriptional responses to environmental inputs. *Molecular Cell*, 37(3):418–428, Feb. 2010.

Gordân, R., Shen, N., Dror, I., Zhou, T., Horton, J., Rohs, R., and Bulyk, M. L. Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell reports*, 3(4):1093–1104, Apr. 2013.

Göttlicher, M. Receptor Toxicology. In Marquardt, H., Schäfer, S., McClellan, R. O., and Welsch, F., editors, *Toxicology*, pages 231–243. Academic Press, 1999.

Grenert, J. P., Sullivan, W. P., Fadden, P., Haystead, T. A., Clark, J., Mimnaugh, E., Krutzsch, H., Ochel, H. J., Schulte, T. W., Sausville, E., Neckers, L. M., and Toft, D. O. The amino-terminal domain of heat shock protein 90 (hsp90) that binds geldanamycin is an ATP/ADP switch domain that regulates hsp90 conformation. *The Journal of biological chemistry*, 272(38):23843–23850, Sept. 1997.

Gu, Y. Z., Hogenesch, J. B., and Bradfield, C. A. The PAS superfamily: sensors of environmental and developmental signals. *Annual review of pharmacology and toxicology*, 40(1):519–561, 2000.

Hahn, M. E., Karchner, S. I., Shapiro, M. A., and Perera, S. A. Molecular evolution of two vertebrate aryl hydrocarbon (dioxin) receptors (AHR1 and AHR2) and the PAS family. *Proceedings of the National Academy of Sciences of the United States of America*, 94(25):13743–13748, Dec. 1997.

Hailfinger, S., Jaworski, M., Braeuning, A., Buchmann, A., and Schwarz, M. Zonal gene expression in murine liver: lessons from tumors. *Hepatology*, 43(3):407–414, Mar. 2006.

Hatzis, P., van der Flier, L. G., van Driel, M. A., Guryev, V., Nielsen, F., Denissov, S., Nijman, I. J., Koster, J., Santo, E. E., Welboren, W., Versteeg, R., Cuppen, E., van de Wetering, M., and Stunnenberg, H. G. Genome-wide pattern of TCF7L2/TCF4 chromatin occupancy in colorectal cancer cells. *Molecular and cellular biology*, 28 (8):2732–2744, Apr. 2008.

He, X., Semenov, M. V., Tamai, K., and Zeng, X. LDL receptor-related proteins 5 and 6 in Wnt/beta-catenin signaling: arrows point the way. *Development*, 131(8): 1663–1677, Apr. 2004.

Heinrich, R. and Neel, B. Mathematical models of protein kinase signal transduction. *Molecular Cell*, 2002.

Heinrich, R. and Schuster, S. *The regulation of cellular systems*. Chapman & Hall, New York, 1996.

Hermsen, R., Ursem, B., and ten Wolde, P. R. Combinatorial gene regulation using auto-regulation. *PLoS computational biology*, 6(6):e1000813, June 2010.

Hill, T. L. *Cooperativity Theory in Biochemistry*. Steady-State and Equilibrium Systems. Springer Science & Business Media, New York, NY, 1985.

Hlavacek, W. S., Faeder, J. R., Blinov, M. L., Posner, R. G., Hucka, M., and Fontana, W. Rules for Modeling Signal-Transduction Systems. *Science Signaling*, 2006(344): re6–re6, July 2006.

Ideker, T., Galitski, T., and Hood, L. A new approach to decoding life: systems biology. *Genomics and Human Genetics*, 2(1):343–372, 2001.

IUPAC. IUPAC-IUB commission on biochemical nomenclature (CBN). Abbreviations and symbols for nucleic acids, polynucleotides and their constituents. *Journal of molecular biology*, 55(3):299–310, Feb. 1971.

Jacob, F., Perrin, D., Sanchez, C., and Monod, J. Operon: a group of genes with the expression coordinated by an operator. *Comptes rendus hebdomadaires des séances de l'Académie des sciences*, 250:1727–1729, Feb. 1960.

Janda, C. Y., Waghray, D., Levin, A. M., Thomas, C., and Garcia, K. C. Structural basis of Wnt recognition by Frizzled. *Science*, 337(6090):59–64, July 2012.

Johnson, A. D., Poteete, A. R., Lauer, G., Sauer, R. T., Ackers, G. K., and Ptashne, M. lambda Repressor and cro–components of an efficient molecular switch. *Nature*, 294(5838):217–223, Nov. 1981.

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K. R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., Palin, K., Vaquerizas, J. M., Vincentelli, R., Luscombe, N. M., Hughes, T. R., Lemaire, P., Ukkonen, E., Kivioja, T., and Taipale, J. DNA-binding specificities of human transcription factors. *Cell*, 152(1-2):327–339, Jan. 2013.

Karlebach, G. and Shamir, R. Modelling and analysis of gene regulatory networks. *Nature Publishing Group*, 9(10):770–780, Oct. 2008.

Kasai, S., Ishigaki, T., Takumi, R., Kamimura, T., and Kikuchi, H. β-catenin signaling induces CYP1A1 expression by disrupting adherens junctions in Caco-2 human colon carcinoma cells. *Biochimica et biophysica acta*, 1830(3):2509–2516, Mar. 2013.

Kazlauskas, A., Poellinger, L., and Pongratz, I. Evidence that the co-chaperone p23 regulates ligand responsiveness of the dioxin (Aryl hydrocarbon) receptor. *The Journal of biological chemistry*, 274(19):13519–13524, May 1999.

Kazlauskas, A., Sundström, S., Poellinger, L., and Pongratz, I. The hsp90 chaperone complex regulates intracellular localization of the dioxin receptor. *Molecular and cellular biology*, 21(7):2594–2607, Apr. 2001.

Kim, H. D. and O'Shea, E. K. A quantitative model of transcription factor-activated gene expression. *Nature Structural and Molecular Biology*, 15(11):1192–1198, Nov. 2008.

Kim, H. D., Shay, T., O'Shea, E. K., and Regev, A. Transcriptional regulatory circuits: predicting numbers from alphabets. *Science*, 325(5939):429–432, July 2009.

Kim, S., Broströmer, E., Xing, D., Jin, J., Chong, S., Ge, H., Wang, S., Gu, C., Yang, L., Gao, Y. Q., Su, X.-d., Sun, Y., and Xie, X. S. Probing allostery through DNA. *Science*, 339(6121):816–819, Feb. 2013.

Kitano, H. Systems biology: a brief overview. *Science*, 295(5560):1662–1664, Mar. 2002a.

Kitano, H. Standards for modeling. *Nature biotechnology*, 20(4):337–337, Apr. 2002b.

Kitano, H. Computational systems biology. *Nature*, 420(6912):206–210, Nov. 2002c.

Klaus, A. and Birchmeier, W. Wnt signalling and its impact on development and cancer. *Nature reviews. Cancer*, 8(5):387–398, May 2008.

Korenčič, A., Bordyugov, G., Košir, R., Rozman, D., Goličnik, M., and Herzel, H. The interplay of cis-regulatory elements rules circadian rhythms in mouse liver. *PLoS ONE*, 7(11):e46835, 2012.

Koss, G. and Wölfle, D. Dioxin and Dioxin-like Polychlorinated Hydrocarbons and Biphenyls. In Marquardt, H., Schäfer, S., McClellan, R. O., and Welsch, F., editors, *Toxicology*, pages 699–711. Academic Press, 1999.

Kress, S., Reichert, J., and Schwarz, M. Functional analysis of the human cytochrome P4501A1 (CYP1A1) gene enhancer. *European journal of biochemistry / FEBS*, 258(2):803–812, Dec. 1998.

Kreutz, C., Raue, A., and Timmer, J. Likelihood based observability analysis and confidence intervals for predictions of dynamic models. *BMC Systems Biology*, 6 (1):120, 2012.

Kubota, M., Kawajiri, K., Sogawa, K., Kaizu, Y., Sawaya, T., Watanabe, J., Gotoh, O., and Fujino, H. Xenobiotic responsive element in the 5'-upstream region of the human P-450c gene. *Journal of biochemistry*, 110(2):232–236, Aug. 1991.

Kuntz, E., Kuntz, E., Kuntz, H.-D., and Kuntz, H.-D. *Hepatology: principles and practice*. History, Morphology, Biochemistry, Diagnostics, Clinic, Therapy. Springer Medizin Verlag, Heidelberg, 2 edition, 2006.

Lam, F. H., Steger, D. J., and O'Shea, E. K. Chromatin decouples promoter threshold from dynamic range. *Nature*, 453(7192):246–250, May 2008.

Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S. H., and Kim, V. N. MicroRNA genes are transcribed by RNA polymerase II. *The EMBO journal*, 23(20):4051–4060, Oct. 2004.

Lehmann, E. L. and Casella, G. *Theory of Point Estimation*. Springer Texts in Statistics. Springer-Verlag, New York, 2 edition, Aug. 1998.

Levchenko, A. Dynamical and integrative cell signaling: challenges for the new biology. *Biotechnology and bioengineering*, 84(7):773–782, Dec. 2003.

Lo, R. and Matthews, J. High-resolution genome-wide mapping of AHR and ARNT binding sites by ChIP-Seq. *Toxicological sciences*, 130(2):349–361, Dec. 2012.

Loeppen, S., Koehle, C., Buchmann, A., and Schwarz, M. A β-catenin-dependent pathway regulates expression of cytochrome P450 isoforms in mouse liver tumors. *Carcinogenesis*, 26(1):239–248, Jan. 2005.

Luch, A. The Carcinogenic Effects of Polycyclic Aromatic Hydrocarbons. Imperial College Press, 2005.

MacDonald, B. T., Tamai, K., and He, X. Wnt/β-catenin signaling: components, mechanisms, and diseases. *Developmental Cell*, 17(1):9–26, July 2009.

MacNeil, L. T. and Walhout, A. J. M. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome research*, 21 (5):645–657, May 2011.

Maerkl, S. J. and Quake, S. R. A systems approach to measuring the binding energy landscapes of transcription factors. *Science*, 315(5809):233–237, Jan. 2007.

Mao, B., Wu, W., Li, Y., Hoppe, D., Stannek, P., Glinka, A., and Niehrs, C. LDL-receptor-related protein 6 is a receptor for Dickkopf proteins. *Nature*, 411(6835): 321–325, May 2001.

Mathelier, A. and Wasserman, W. W. The next generation of transcription factor binding site prediction. *PLoS computational biology*, 9(9):e1003214, 2013.

Mathelier, A., Zhao, X., Zhang, A. W., Parcy, F., Worsley-Hunt, R., Arenillas, D. J., Buchman, S., Chen, C.-y., Chou, A., Ienasescu, H., Lim, J., Shyr, C., Tan, G., Zhou, M., Lenhard, B., Sandelin, A., and Wasserman, W. W. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Research*, 42(Database issue):D142–7, Jan. 2014.

Mathew, L. K., Sengupta, S. S., Ladu, J., Andreasen, E. A., and Tanguay, R. L. Crosstalk between AHR and Wnt signaling through R-Spondin1 impairs tissue regeneration in zebrafish. *FASEB journal*, 22(8):3087–3096, Aug. 2008.

Matys, V., Kel-Margoulis, O. V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., Voss, N., Stegmaier, P., Lewicki-Potapov, B., Saxel, H., Kel, A. E., and Wingender, E. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Research*, 34(Database issue):D108–10, Jan. 2006.

Mayo, A. E., Setty, Y., Shavit, S., Zaslaver, A., and Alon, U. Plasticity of the cis-regulatory input function of a gene. *PLoS Biology*, 4(4):e45, Apr. 2006.

McGuire, J., Whitelaw, M. L., Pongratz, I., Gustafsson, J. A., and Poellinger, L. A cellular factor stimulates ligand-dependent release of hsp90 from the basic helix-loop-helix dioxin receptor. *Molecular and cellular biology*, 14(4):2438–2446, Apr. 1994.

McKay, M. D., Beckman, R. J., and Conover, W. J. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, 21(2):239–245, May 1979.

Meeker, W. Q. and Escobar, L. A. Teaching about Approximate Confidence Regions Based on Maximum Likelihood Estimation. *The American Statistician*, 49(1):48–53, Feb. 1995.

Meyer, B. K. and Perdew, G. H. Characterization of the AhR-hsp90-XAP2 core complex and the role of the immunophilin-related protein XAP2 in AhR stabilization. *Biochemistry*, 38(28):8907–8917, July 1999.

Meyer, B. K., Pray-Grant, M. G., Vanden Heuvel, J. P., and Perdew, G. H. Hepatitis B virus X-associated protein 2 is a subunit of the unliganded aryl hydrocarbon receptor core complex and exhibits transcriptional enhancer activity. *Molecular and cellular biology*, 18(2):978–988, Feb. 1998.

Miller, J. C. and Pabo, C. O. Rearrangement of side-chains in a Zif268 mutant highlights the complexities of zinc finger-DNA recognition. *Journal of molecular biology*, 313(2):309–315, Oct. 2001.

Molenaar, M., van de Wetering, M., Oosterwegel, M., Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., and Destrée, O. XTcf-3 transcription factor mediates β-catenin-induced axis formation in Xenopus embryos. *Cell*, 86(3):391–399, Aug. 1996.

Monod, J., Wyman, J., and Changeux, J. P. On the Nature of Allosteric Transitions: A Plausible Model. *Journal of molecular biology*, 12:88–118, May 1965.

Nair, S. C., Toran, E. J., Rimerman, R. A., Hjermstad, S., Smithgall, T. E., and Smith, D. F. A pathway of multi-chaperone interactions common to diverse regulatory proteins: estrogen receptor, Fes tyrosine kinase, heat shock transcription factor Hsf1, and the aryl hydrocarbon receptor. *Cell stress & chaperones*, 1(4):237–250, Dec. 1996.

Niehrs, C. The complex world of WNT receptor signalling. *Nature Publishing Group*, 13(12):767–779, Dec. 2012.

Nishida, K., Frith, M. C., and Nakai, K. Pseudocounts for transcription factor binding sites. *Nucleic Acids Research*, 37(3):939–944, Feb. 2009.

Oesch, F. and Arand, M. Xenobiotic Metabolism. In Marquardt, H., Schäfer, S., McClellan, R. O., and Welsch, F., editors, *Toxicology*, pages 83–108. Academic Press, 1999.

Okino, S. T. and Whitlock, J. P. Dioxin induces localized, graded changes in chromatin structure: implications for Cyp1A1 gene transcription. *Molecular and cellular biology*, 15(7):3714–3721, July 1995.

Paul, S., Vadgama, P., and Ray, A. K. Surface plasmon resonance imaging for biosensing. *IET nanobiotechnology*, 3(3):71–80, Sept. 2009.

Perdew, G. H. Association of the Ah receptor with the 90-kDa heat shock protein. *The Journal of biological chemistry*, 263(27):13802–13805, Sept. 1988.

Perdew, G. H. and Bradfield, C. A. Mapping the 90 kDa heat shock protein binding region of the Ah receptor. *Biochemistry and molecular biology international*, 39(3): 589–593, June 1996.

Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*, 29(9):e45, May 2001.

Pinson, K. I., Brennan, J., Monkley, S., Avery, B. J., and Skarnes, W. C. An LDL-receptor-related protein mediates Wnt signalling in mice. *Nature*, 407(6803):535–538, Sept. 2000.

Polakis, P. Wnt signaling and cancer. *Genes & development*, 14(15):1837–1851, Aug. 2000.

Polakis, P. The many ways of Wnt in cancer. *Current opinion in genetics & development*, 17(1):45–51, Feb. 2007.

Pongratz, I., Mason, G. G., and Poellinger, L. Dual roles of the 90-kDa heat shock protein hsp90 in modulating functional activities of the dioxin receptor. *The Journal of biological chemistry*, 267(19):13728–13734, July 1992.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. Numerical Recipes in Fortran 77. Cambridge University Press, 1992.

Ptashne, M. and Gann, A. *Genes & signals*. Cold Spring Harbor, New York : Cold Spring Harbor Laboratory Press, 2002.

Puga, A., Ma, C., and Marlowe, J. L. The aryl hydrocarbon receptor cross-talks with multiple signal transduction pathways. *Biochemical pharmacology*, 77(4):713–722, Feb. 2009.

Rahmann, S., Müller, T., and Vingron, M. On the power of profiles for transcription factor binding site detection. *Statistical applications in genetics and molecular biology*, 2:Article7, 2003.

Ramsey, S. A., Klemm, S. L., Zak, D. E., Kennedy, K. A., Thorsson, V., Li, B., Gilchrist, M., Gold, E. S., Johnson, C. D., Litvak, V., Navarro, G., Roach, J. C., Rosenberger, C. M., Rust, A. G., Yudkovsky, N., Aderem, A., and Shmulevich, I. Uncovering a macrophage transcriptional program by integrating evidence from motif scanning and expression dynamics. *PLoS computational biology*, 4(3): e1000021, Mar. 2008.

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmuller, U., and Timmer, J. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25 (15):1923–1929, Aug. 2009.

Raue, A., Kreutz, C., Maiwald, T., Klingmüller, U., and Timmer, J. Addressing parameter identifiability by model-based experimentation. *IET systems biology*, 5 (1):120–130, Feb. 2011.

Reya, T. and Clevers, H. Wnt signalling in stem cells and cancer. *Nature*, 434(7035): 843–850, Apr. 2005.

Roose, J., Molenaar, M., Peterson, J., Hurenkamp, J., Brantjes, H., Moerer, P., van de Wetering, M., and Destrée, O. The Xenopus Wnt effector XTcf-3 interacts with Groucho-related transcriptional repressors. *Nature*, 395(6702):608–612, Oct. 1998.

Saez-Rodriguez, J., Alexopoulos, L. G., Epperlein, J., Samaga, R., Lauffenburger, D. A., Klamt, S., and Sorger, P. K. Discrete logic modelling as a means to link

protein signalling networks with functional analysis of mammalian signal transduction. *Molecular systems biology*, 5(1):331, 2009.

Schliess, F., Hoehme, S., Henkel, S. G., Ghallab, A., Driesch, D., Böttger, J., Guthke, R., Pfaff, M., Hengstler, J. G., Gebhardt, R., Häussinger, D., Drasdo, D., and Zellmer, S. Integrated metabolic spatial-temporal model for the prediction of ammonia detoxification during liver damage and regeneration. *Hepatology*, 60(6): 2040–2051, May 2014.

Schmidt, J. V. and Bradfield, C. A. Ah receptor signaling pathways. *Annual review of cell and developmental biology*, 12(1):55–89, 1996.

Schneider, A. J., Branam, A. M., and Peterson, R. E. Intersection of AHR and Wnt signaling in development, health, and disease. *International journal of molecular sciences*, 15(10):17852–17885, 2014.

Schneider, T. D. and Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Research*, 18(20):6097–6100, Oct. 1990.

Schreiber, T. D., Köhle, C., Buckler, F., Schmohl, S., Braeuning, A., Schmiechen, A., Schwarz, M., and Münzel, P. A. Regulation of CYP1A1 gene expression by the antioxidant tert-butylhydroquinone. *Drug metabolism and disposition: the biological fate of chemicals*, 34(7):1096–1101, July 2006.

Schulthess, P. and Blüthgen, N. From reaction networks to information flow–using modular response analysis to track information in signaling networks. *Methods in enzymology*, 500:397–409, Dec. 2010.

Schulthess, P., Löffler, A., Vetter, S., Kreft, L., Schwarz, M., Braeuning, A., and Blüthgen, N. Signal integration by the CYP1A1 promoter - a quantitative study. *Nucleic Acids Research*, 43(11):5318–5330, June 2015.

Schwarz, L. and Watkins, J. B. The Liver. In Greim, H. and Snyder, R., editors, *Toxicology and Risk Assessment*. Wiley-Interscience, Chichester, UK, Jan. 2008.

Schwen, L. O., Krauss, M., Niederalt, C., Gremse, F., Kiessling, F., Schenk, A., Preusser, T., and Kuepfer, L. Spatio-Temporal Simulation of First Pass Drug Perfusion in the Liver. *PLoS computational biology*, 10(3):e1003499, Mar. 2014.

Schwenk, M. The Gastrointestinal Tract. In Greim, H. and Snyder, R., editors, *Toxicology and Risk Assessment*. Wiley-Interscience, Chichester, UK, Jan. 2008.

Seber, G. A. F. and Wild, C. J. *Nonlinear Regression*. John Wiley & Sons, Sept. 2003.

Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D., and Friedman, N. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nature genetics*, 34(2):166–176, June 2003.

Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U., and Gaul, U. Predicting expression patterns from regulatory sequence in Drosophila segmentation. *Nature*, 451(7178):535–540, Jan. 2008.

Setty, Y., Mayo, A. E., Surette, M. G., and Alon, U. Detailed map of a cis-regulatory input function. *Proceedings of the National Academy of Sciences of the United States of America*, 100(13):7702–7707, June 2003.

Shea, M. A. and Ackers, G. K. The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation. *Journal of molecular biology*, 181(2): 211–230, Jan. 1985.

Sherman, M. S. and Cohen, B. A. Thermodynamic State Ensemble Models of cis-Regulation. *PLoS computational biology*, 8(3):e1002407, Mar. 2012.

Shumaker-Parry, J. S., Aebersold, R., and Campbell, C. T. Parallel, quantitative measurement of protein binding to a 120-element double-stranded DNA array in real time using surface plasmon resonance microscopy. *Analytical chemistry*, 76 (7):2071–2082, Apr. 2004.

Siggers, T. and Gordân, R. Protein-DNA binding: complexities and multi-protein codes. *Nucleic Acids Research*, 42(4):2099–2111, Feb. 2014.

Siggers, T., Chang, A. B., Teixeira, A., Wong, D., Williams, K. J., Ahmed, B., Ragoussis, J., Udalova, I. A., Smale, S. T., and Bulyk, M. L. Principles of dimer-specific gene regulation revealed by a comprehensive characterization of NF-κB family DNA binding. *Nature immunology*, 13(1):95–102, Jan. 2012.

Skehan, P., Storeng, R., Scudiero, D., Monks, A., McMahon, J., Vistica, D., Warren, J. T., Bokesch, H., Kenney, S., and Boyd, M. R. New colorimetric cytotoxicity

assay for anticancer-drug screening. *Journal of the National Cancer Institute*, 82(13): 1107–1112, July 1990.

Spitz, F. and Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nature reviews. Genetics*, 13(9):613–626, Sept. 2012.

Staden, R. Computer methods to locate signals in nucleic acid sequences. *Nucleic Acids Research*, 12(1 Pt 2):505–519, Jan. 1984.

Stelniec-Klotz, I., Legewie, S., Tchernitsa, O., Witzel, F., Klinger, B., Sers, C., Herzel, H., Blüthgen, N., and Schäfer, R. Reverse engineering a hierarchical regulatory network downstream of oncogenic KRAS. *Molecular systems biology*, 8(1):601, 2012.

Stockinger, B., Di Meglio, P., Gialitakis, M., and Duarte, J. H. The aryl hydrocarbon receptor: multitasking in the immune system. *Annual Review of Immunology*, 32 (1):403–432, 2014.

Stormo, G. D. DNA binding sites: representation and discovery. *Bioinformatics*, 16 (1):16–23, Jan. 2000.

Stormo, G. D. and Hartzell, G. W. Identifying protein-binding sites from unaligned DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America*, 86(4):1183–1187, Feb. 1989.

Stormo, G. D. and Zhao, Y. Determining the specificity of protein-DNA interactions. *Nature reviews. Genetics*, 11(11):751–760, Nov. 2010.

Stormo, G. D., Schneider, T. D., Gold, L., and Ehrenfeucht, A. Use of the 'Perceptron' algorithm to distinguish translational initiation sites in E. coli. *Nucleic Acids Research*, 10(9):2997–3011, May 1982.

Sugioka, K., Mizumoto, K., and Sawa, H. Wnt regulates spindle asymmetry to generate asymmetric nuclear β-catenin in C. elegans. *Cell*, 146(6):942–954, Sept. 2011.

Swanson, H. I., Chan, W. K., and Bradfield, C. A. DNA binding specificities and pairing rules of the Ah receptor, ARNT, and SIM proteins. *The Journal of biological chemistry*, 270(44):26292–26302, Nov. 1995.

Tamai, K., Semenov, M. V., Kato, Y., Spokony, R., Liu, C., Katsuyama, Y., Hess, F., Saint-Jeannet, J. P., and He, X. LDL-receptor-related proteins in Wnt signal transduction. *Nature*, 407(6803):530–535, Sept. 2000.

Tamai, K., Zeng, X., Liu, C., Zhang, X., Harada, Y., Chang, Z., and He, X. A mechanism for Wnt coreceptor activation. *Molecular Cell*, 13(1):149–156, Jan. 2004.

Thiel, C., Schneckener, S., Krauss, M., Ghallab, A., Hofmann, U., Kanacher, T., Zellmer, S., Gebhardt, R., Hengstler, J. G., and Kuepfer, L. A systematic evaluation of the use of physiologically based pharmacokinetic modeling for cross-species extrapolation. *Journal of pharmaceutical sciences*, 104(1):191–206, Jan. 2015.

Tomovic, A. and Oakeley, E. J. Position dependencies in transcription factor binding sites. *Bioinformatics*, 23(8):933–941, Apr. 2007.

Torre, C., Perret, C., and Colnot, S. Molecular determinants of liver zonation. *Progress in molecular biology and translational science*, 97:127–150, 2010.

Vaas, S., Kreft, L., Schwarz, M., and Braeuning, A. Cooperation of structurally different aryl hydrocarbon receptor agonists and β-catenin in the regulation of CYP1A expression. *Toxicology*, 325:31–41, Nov. 2014.

Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F., and Jaenisch, R. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*, 153(4):910–918, May 2013.

Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T. W., Greven, M. C., Pierce, B. G., Dong, X., Kundaje, A., Cheng, Y., Rando, O. J., Birney, E., Myers, R. M., Noble, W. S., Snyder, M., and Weng, Z. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome research*, 22(9):1798–1812, Sept. 2012.

Wang, S. and Hankinson, O. Functional involvement of the Brahma/SWI2-related gene 1 protein in cytochrome P4501A1 transcription mediated by the aryl hydrocarbon receptor complex. *The Journal of biological chemistry*, 277(14):11821–11827, Apr. 2002.

Wehrli, M., Dougan, S. T., Caldwell, K., O'Keefe, L., Schwartz, S., Vaizel-Ohayon, D., Schejter, E., Tomlinson, A., and DiNardo, S. arrow encodes an LDL-receptor-related protein essential for Wingless signalling. *Nature*, 407(6803):527–530, Sept. 2000.

Weirauch, M. T., Cote, A., Norel, R., Annala, M., Zhao, Y., Riley, T. R., Saez-Rodriguez, J., Cokelaer, T., Vedenko, A., Talukder, S., DREAM5 Consortium, Bussemaker, H. J., Morris, Q. D., Bulyk, M. L., Stolovitzky, G., and Hughes, T. R. Evaluation of methods for modeling transcription factor sequence specificity. *Nature biotechnology*, 31(2):126–134, Feb. 2013.

Whitlock, J. P., Okino, S. T., Dong, L., Ko, H. P., Clarke-Katzenberg, R., Ma, Q., and Li, H. Cytochromes P450 5: induction of cytochrome P4501A1: a model for analyzing mammalian gene transcription. *FASEB journal*, 10(8):809–818, June 1996.

Wong, D., Teixeira, A., Oikonomopoulos, S., Humburg, P., Lone, I. N., Saliba, D., Siggers, T., Bulyk, M., Angelov, D., Dimitrov, S., Udalova, I. A., and Ragoussis, J. Extensive characterization of NF-κB binding uncovers non-canonical motifs and advances the interpretation of genetic functional traits. *Genome biology*, 12(7):R70, 2011.

Yao, E. F. and Denison, M. S. DNA sequence determinants for binding of transformed Ah receptor to a dioxin-responsive enhancer. *Biochemistry*, 31(21):5060–5067, June 1992.

Yuh, C. H., Bolouri, H., and Davidson, E. H. Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene. *Science*, 279(5358): 1896–1902, Mar. 1998.

Zenklusen, D., Larson, D. R., and Singer, R. H. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nature Structural and Molecular Biology*, 15(12):1263–1271, Dec. 2008.

Zhou, T., Yang, L., Lu, Y., Dror, I., Dantas Machado, A. C., Ghane, T., Di Felice, R., and Rohs, R. DNAshape: a method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic Acids Research*, 41(Web Server issue):W56–62, July 2013.

List of Figures

List of Tables

List of Abbreviations

3-MC 3-methylcholanthrene. 3, 24, 74, 75

AhR aryl hydrocarbon receptor. 7–9, 20–24, 41–51, 53–56, 62, 64–66, 71–75, 77, 78, 80–83, 122

AIP AhR-interacting protein. 7, 8

APC adenomatous polyposis coli. 9

Arnt AhR nuclear translocator. 8, 44, 47–50, 54–56, 62, 77, 78, 80

AXIN1 axis inhibitor 1. 9, 10

CK1α/δ casein kinase 1α/δ. 9

CK1γ casein kinase 1γ. 10

CYP cytochrome P450. 4, 5, 7, 77

CYP1A1 cytochrome P450, family 1, subfamily A, polypetide 1. v–viii, 3, 5, 6, 8, 10, 20–22, 33, 41–45, 47, 49–51, 53–55, 66–68, 71, 73–79, 81–83, 85

CYP1A2 cytochrome P450, family 1, subfamily A, polypetide 2. 5

DRE dioxin responsive element. v, 5, 6, 8, 12, 21, 22, 42, 44, 48–54, 56–69, 71, 73, 74, 78–83, 90–92, 95, 96

DVL Dishevelled. 10

ER endoplasmic reticulum. 4, 5

FZD Frizzled. 9, 10

GS glutamine synthetase. 83

GSK3α/β glycogen synthase kinase $3\alpha/\beta$. 9, 10

HNF1 hepatocyte nuclear factor 1. 5

HNF3 hepatocyte nuclear factor 3. 5

HNF4 hepatocyte nuclear factor 4. 5

HSP90 heat shock protein 90. 7, 8

LRP5/6 low-density lipoprotein receptor-related protein 5 or 6. 9, 10

MWC Monod-Wyman-Changeur. 16

PSCM position specific count matrix. 28–30, 38, 49

PWM position specific weight matrix. 14, 15, 31, 47, 50, 51

RNAP RNA polymerase. v, 11, 16–19, 33–35, 57–59, 61, 62, 64, 69, 70, 80–82, 89–92

TCDD 2,3,7,8-Tetrachlorodibenzo-p-dioxin. v, vii, 1–3, 7, 8, 22, 24, 42–46, 51–56, 59–62, 64–68, 71–75, 77, 79–81, 95, 96

TCF4 transcription factor 4, also known as TCF7L2. 48–50

TCF7L2 transcription factor 7-like 2. 6, 10, 21, 47–50, 126

TF transcription factor. v, vi, 6–17, 19, 20, 30, 32–35, 41, 44–58, 60–64, 68–70, 77–82, 85, 86, 89

TFBS transcription factor binding site. 5, 13–15, 21, 33, 35, 41, 42, 44, 47, 50–54, 57, 61–65, 67–69, 71, 79, 80, 82, 90, 91, 95

XRE xenobiotic response element. 5

List of Publications

Schulthess, P., Löffler, A., Vetter, S., Kreft, L., Schwarz, M., Braeuning, A., and Blüthgen, N.. Signal integration by the *CYP1A1* promoter — a quantitative study. *Nucleic Acids Research*, 43(11):5318–5330, 2015

Thomas, P., Durek, P., Solt, I., Klinger, B., Witzel, F., **Schulthess, P.**, Mayer, Y., Tikk, D., Blüthgen, N., and Leser, U.. Computer-assisted curation of a human regulatory core network from the biological literature. *Bioinformatics*, 31(8):1258–1266, 2015

Schulthess, P. and Blüthgen, N.. From reaction networks to information flow — using modular response analysis to track information in signaling networks. *Methods in enzymology*, 500:397–409, 2010

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel erarbeitet und verfasst habe. Ich besitze keinen entsprechenden Doktorgrad und habe mich nicht anderwärts um einen Doktorgrad beworben. Die dem Promotionsverfahren zugrunde liegende Promotionsordnung ist mir bekannt.

Berlin, den 29. Oktober 2015 Pascal Schulthess